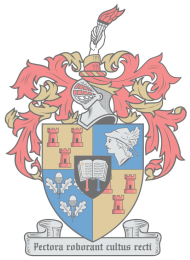
A survey of the YAN status of South African grape juices and exploration of multivariate data analysis techniques for spectrometric calibration and cultivar discrimination purposes

by

Gabriella Petrovic



UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

Thesis presented in partial fulfilment of the requirements for the degree of
Master of Science

at
Stellenbosch University
Institute for Wine Biotechnology, Faculty of AgriSciences

Supervisor: Dr Astrid Buica
Co-supervisor: Dr José Luis Aleixandre-Tudo

December 2018

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2018

Summary

Yeast assimilable nitrogen (YAN) has been identified as one of the main drivers of wine quality, influencing the production of various aromas and ensuring a successful fermentation to dryness. Due to the number of factors affecting YAN concentration and composition, paired with the complexities of yeast metabolism, more data is required to enable a comprehensive understanding of this important component of the grape juice matrix. Thus, there is a need for simple, rapid, and cost-effective methods to measure YAN status. The main aims of this research were to gain insight into the nitrogen status of grape juices used for commercial winemaking in the South African wine industry, and subsequently, to assist in a more comprehensive understanding of grape juice nitrogen status.

Therefore, in Chapter 3, an unsupervised survey of the YAN, FAN, and ammonia concentrations of 805 grape juice samples of various (industrially relevant) cultivars and geographical origins are reported. Subsequently, an overall average of 191 ± 64 mg N/L, 138 ± 46 mg N/L and 53 ± 24 mg N/L was observed for YAN, FAN, and ammonia, respectively. Trends of nitrogen deficiency and excess could be found for various cultivars and geographical origins. Analysis of variance tests and exploratory data analysis techniques such as hierarchical agglomerative clustering and CART analysis established ‘cultivar’ as the most important factor in determining the YAN concentration and composition of the resulting grape juice.

In Chapter 4, using the data collected in Chapter 3, plus an additional vintage (2018), the viability of infrared (IR) spectroscopy for the accurate quantification of YAN, FAN, and ammonia was tested. IR spectroscopies compared included: Fourier-transform infrared (FT-IR), Fourier-transform near infrared (FT-NIR) and attenuated total reflection mid-infrared (ATR-MIR) spectroscopy. FT-IR and FT-NIR were found to outperform ATR-MIR in a variety of tasks assigned to each instrument and were deemed robust and capable of accurate quantification as RPD_{VAL} > 2.5 were repeatedly obtained for both spectroscopies. The achievement of accurate calibration models is owed to the large amount of variability included in both the calibration and validation sets and the application of proper external validation strategies. Thus, both industry and research are presented with a simple, rapid and cost-effective method to measure this important component of the grape juice matrix.

In Chapter 5, a deeper look into the FAN component of YAN was conducted by quantifying individual amino acids. Overall, proline, arginine, glutamine, alanine, tryptophan and GABA were found to be the most abundant while glycine, lysine, methionine and, ornithine were found to be the least abundant. Subsequently, the discriminatory power of the amino acid profile of the various cultivars were tested. This was done to identify key differences in amino acid profiles which could possibly serve as the basis for further research investigating yeast metabolism and aroma production during fermentation.

The results of this research have contributed a wealth of information regarding the nitrogen status of various cultivars of *Vitis vinifera*, together with a rapid and easy-to-use method for the quantification of the nitrogen status of the grape juice matrix. This was done in hope of furthering the research efforts in this field to aid the production of quality wines, capable of meeting consumer demands.

Opsomming

Gis-aanneembare stikstof (YAN) is geïdentifiseer as een van die belangrikste dryfkragte van wynkwaliteit, wat die produksie van verskillende aromas beïnvloed en 'n suksesvolle fermentasie-tot-droogte verseker. As gevolg van die aantal faktore wat die konsentrasie en die samestelling van YAN beïnvloed, tesame met die kompleksiteite van gismetabolisme, word meer data benodig om 'n omvattende begrip van hierdie belangrike komponent van die druiwesapmatrikste te bewerkstellig. Daar is dus 'n behoefte aan eenvoudige, vinnige en koste-effektiewe metodes om YAN-status te meet. Die hoofdoelwitte van hierdie navorsing was om insig te verkry van die stikstofstatus van druiwesap, wat gebruik word vir kommersiële wynmaak in die Suid-Afrikaanse wynbedryf. Die studie beoog verder om 'n meer omvattende begrip van druiwesap-stikstofstatus ten toon te stel.

Hoofstuk 3 lewer resultate oor die opname van die YAN-, FAN- en ammoniak-konsentrasies van 805 druiwesapmonsters van verskeie kultivars en geografiese distrikte. 'n Algehele gemiddeld van 191 ± 64 mg N/L, 138 ± 46 mg N/L en 53 ± 24 mg N/L is waargeneem vir onderskeidelik YAN, FAN en ammoniak. Neigings van stikstof-tekorte en -oormaat is gevind vir verskillende kultivars en verskillende geografiese distrikte. Analise van variansie toetse en ‘n verskeidenheid data analise tegnieke soos hiërargiese agglomeratiewe ‘clustering’ en CART analise het 'kultivar' as die belangrikste faktor in die bepaling van die YAN konsentrasie en samestelling van druiwesap bevestig.

In hoofstuk 4 word die data van hoofstuk 3 gekombineer met ‘n addisionele oesjaar (2018), om die geskiktheid van infrarooi (IR) spektroskopie vir die akkurate kwantifisering van YAN, FAN en ammoniak te toets. Die IR-spektroskopieë wat vergelyk is sluit in: Fourier-transform infrarooi (FT-IR), Fourier-transform nabye infrarooi (FT-NIR) en verswakte totale refleksie mid-infrarooi (ATR-MIR) spektroskopie. Daar is gevind dat FT-IR en FT-NIR herhaaldelik beter presteer as ATR-MIR in 'n verskeidenheid take – vir beide FT-IR en FT-NIR is RPD_{VAL}> 2.5 konstant verkry. Die akkuraatheid van die kalibrasie modelle kan toegeskryf word aan die groot hoeveelheid veranderlikes wat ingesluit is in beide die kalibrasie- en valideringsstelle, tesame met die toepassing van behoorlike eksterne valideringstrategieë. Dus, die modelle bied 'n eenvoudige, vinnige en koste-effektiewe metode aan die industrie en die akademie om hierdie belangrike komponent van die druiwesapmatriks te meet.
In Hoofstuk 5 is 'n meer in-diepte ondersoek na die FAN-komponent van YAN gedoen deur individuele aminosure te kwantifiseer. Daar is grotendeels bevind dat proline, arginien, glutamien, alanien, tryptofaan en GABA in die hoogste konsentrasies voorkom; terwyl glikien, lysien, metionien en ornitien in die laagste konsentrasies voorkom. Vervolgens is die onderskeid in die aminosuurprofiel gebruik om die verskillende kultivars te identifiseer. Dit is gedoen om sleutel verskille in aminosuurprofiele te identifiseer wat moontlik kan dien as die basis vir verdere navorsing van gismetabolisme en aroma-produksie tydens fermentasie.

Die resultate van hierdie navorsing het tot 'n wye basis van inligting bygedra rakende die stikstofstatus van verskeie kultivars van *Vitis vinifera*, sowel as 'n vinnige en maklike metode om die stikstofstatus van die druiwesapmatriks te kwantifiseer. Dié studie is gedoen om die navorsingspogings in hierdie veld te bevorder en die produksie van gehalte wyne te bereik wat sal voldoen aan verbruikersbehoeftes.

This thesis is dedicated to my late father whose passion for life and knowledge I carry with me every day.

Biographical sketch

Gabriella Petrovic was born in Roodepoort, Johannesburg on 19 March 1993. After her family relocated to the Western Cape in 1996, she attended La Rochelle Pre-primary and in 2000, started her primary school career at Gordon's Bay Primary. In 2007 she went on to attend Parel Vallei High School and in 2012, enrolled for a BSc-degree in Molecular Biology and Biotechnology at the Stellenbosch University. Her passion for the wine industry spurred on her postgraduate studies at the Institute for Wine Biotechnology where she enrolled for a HonsBSc-degree in 2016 and an MSc-degree in 2017.

Acknowledgements

I wish to express my sincere gratitude and appreciation to the following persons and institutions:

- My supervisor, **Dr Astrid Buica** for her unwavering support, commitment, guidance, and enthusiasm throughout this process. Thank you for not only mentoring me in science but in life.
- My co-supervisor, **Dr José-Luis Aleixandre Tudo** for his passion for research, his critical thinking and his kind heart.
- **Professor Martin Kidd** for his time and assisting my discovery of the world of statistics.
- **Mr Malcolm Taylor, Mr Erick van Schalkwyk, and Dr Marietjie Stander** for not only the help with chemical analyses and patience to answer all my questions but also for the long chats and much welcomed sarcasm.
- **The winemakers** for greeting me with a smile every week during harvest and if I'm lucky, a cup of coffee and a chat.
- **Professor Benoit Divol** for his open door and accepting me as an unofficial member of his lab. Thank you for always being up for a chat.
- **Karin Vergeer** for her administrative assistance and kind nature.
- **Winetech and NRF** for funding this research.
- My boyfriend, **Berno Greyling** for teaching me the art of perseverance and supporting me every step of the way. Every day your gracious and diligent nature inspires me to be a better version of myself.
- My dear friend **Dr Stephanie Rollero** who made every day at the department a better place. Thank you for putting up with my drama-queen antics, being my soundboard for scientific reasoning and understanding that some days, you just need to sit on the lab floor.
- My dear friend **Claudia Gevers** for the sunshine she brought into my life in the days writing up this thesis. Your resilience and positivity is truly inspiring.
- My dear friend **Carla Snyman** for being there through the good, the bad, and the ugly and for her contagious laughter.
- My loving **father, Peter Petrovic,** for making all this possible. Thank you for all your wisdom, support, and love that still carries me through every day.

Preface

This thesis is presented as a compilation of 6 chapters.

Chapter 1	**Background and research aims and objectives**
Chapter 2	**Literature review** Unravelling the complexities of wine: A Big Data approach.
Chapter 3	**Research results** A statistical exploration of survey data to identify the role of cultivar and origin in the concentration and composition of yeast assimilable nitrogen.
Chapter 4	**Research results** Viability of IR spectroscopy for the accurate measurement of N content of grape juice.
Chapter 5	**Research results** Grape must profiling and cultivar discrimination based on amino acid composition.
Chapter 6	**General discussion and conclusions**

Table of Contents

Chapter 1

Background and Research Aims and Objectives

Chapter 1

Background and Research Aims and Objectives

1.1. Introduction

Never has there been, nor will there ever be such an enigmatic and mystical beverage as wine. This beverage, often referred to as "*poetry in a bottle*" has long been regarded as an artform rather than a scientific process (Swiegers *et al.*, 2005). Over the centuries, winemaking has followed an empirical approach, with the 'know-how' being passed down from generation to generation. Thus, due to the romanticism associated with wine and the belief that wine is an expression of place (hence the French term '*terroir'),* there has traditionally been a minimalistic approach to winemaking. As such, the grapes, and subsequently the wine, should reflect the environment that it was grown in (Bisson *et al.*, 2002). However, as wine has become an increasingly important commodity worldwide – with 24.67 billion litres produced globally in 2017 (Decanter News, 2017) – the development of innovative technologies is becoming a crucial factor in determining the success of the industry as a whole (Pretorius & Høj, 2005). Due to the deep-rooted and rich history of wine, compared to other industries, the acceptance of technological advances and innovation has been slow (Bisson *et al.*, 2002; Pretorius & Høj, 2005; Dambergs *et al.*, 2015).

Unlike other fermented beverages, wine's appeal does not stem from its consistent flavour and aroma, but instead from the unique sensory experience that it can offer, from one vintage to the next (Bisson *et al.*, 2002). Thus, the consistency that is expected from a particular producer refers to a certain level of *quality* rather than a consistency in flavour. However, what constitutes as a 'quality wine' has become an increasingly controversial subject of debate (Pettigrew & Charters, 2006; Johnson & Bruwer, 2007; Lockshin & Corsi, 2012). The emphasis that has been placed on understanding the conundrum of wine quality is primarily due to the increasing power of consumers – fuelled by globalisation and the free flow of information. As a result, the world-wine market has been forced to adopt a more *market-driven* approach (Pretorius & Høj, 2005). A few decades ago, quality was the prerogative of the producer, and consumers who did not appreciate a certain style of wine were often regarded as uncultured by their more affluent counterparts (Bisson *et al.*, 2002). However, in modern times, the definition of quality has moved into the hands of the consumer (Hui, 2006), and at the same time, has become a more subjective concept (Pretorius & Høj, 2005).

Moreover, this paradigm shift of power between the producer and the consumer has put, particularly the 'Old-World' wine producer, in a vulnerable position. Aside from the dwindling wine consumption patterns reported for the Old-world (in contrast to what is being observed in the 'New-World') (Campbell & Guibert, 2006), the advantage that these emerging wine producers have is said to be

their willingness to implement innovative technologies, which, in some cases, may involve stepping away from tradition (Cunico, 2014). As expressed in the words of a New-World wine producer:

> *"The mystery and the magic and the human element do not need to decrease because of the presence of technology. These things are going to happen with or without you and you'll just spend your time wondering what happened." – Palmaz Family Winery, Napa Valley, California.*

It must, however, be made clear that the technology being referred to here is over and above the tools required for the success of large-scale production. In other words, instead of mechanisation and the general protocols ubiquitously employed to reduce the chances of spoilage, innovation in the wine industry is shifting towards optimizing *quality* (Pretorius & Høj, 2005). This is especially relevant as current consumer trends are indicating a preference for more premium wines (Bernetti *et al.*, 2006; Fine Wine Report 2018).

As the grape juice matrix provides the nutrients required by the yeast during fermentation, it has been identified as the primary determinant of the quality of the final wine (Fleet, 2003). Consequently, to satisfy the needs of the modern-day consumer, an in-depth investigation into the chemical constituents present in the grape juice matrix has been conducted. Over the years, through these investigations, yeast assimilable nitrogen (YAN) has been identified as one of the key role-players (Bell & Henschke, 2005).

YAN, primarily constituted by α-amino nitrogen and ammonium ions, is crucial for the growth of the yeast, and subsequently, ensuring fermentation to dryness (Henschke & Jiranek, 1993; Bell & Henschke, 2005; Torrea *et al.*, 2011). This has been confirmed through numerous studies which have identified a deficiency in easily assimilable nitrogen as the main cause for stuck/sluggish fermentations (Bely *et al.*, 1990a,b; Bisson, 1999). Furthermore, in addition to fulfilling the biosynthetic activity of the yeast, YAN has been observed to have a major impact on the production of various volatile (higher alcohols, esters, and volatile fatty acids) and non-volatile metabolites (glycerol, succinic acid, malic acid, and α-ketoglutaric acid), influencing the organoleptic properties, and subsequently, the quality of the final wine (Vilanova *et al.*, 2007; Torrea *et al.*, 2011; Rollero *et al.*, 2018). However, as this research field progresses, the *complexity* of the mechanisms by which YAN leads to the formation of these various aroma compounds is becoming increasingly more evident.

1.2. Research Aims and Objectives

Taking this into consideration, there is a need to understand, predict and monitor the YAN concentration and composition of the grape juice matrix. Thus, developing and adopting technologies which will enable the easy monitoring and control of this important component of the grape juice

matrix can facilitate more informed decision-making, and thus, can increase the chances of producing premium wines. Therefore, the research aims, and objectives of this thesis are as follows:

Aim 1: Gain insight into the YAN status of South African grape juices currently used for commercial winemaking.

- Conduct a survey over the course of two vintages by collecting, analysing, and reporting on the YAN concentration and composition of various industrially relevant cultivars, originating from an array of growing districts stretching across the Western Cape region of South Africa.

Aim 2: Help provide a more comprehensive understanding of the YAN status of the grape juice matrix.

- Maximise the information output of the surveyed data using various descriptive and exploratory statistical techniques.
- Building robust quantitative models (using IR spectroscopy) for the measurement of total YAN, FAN, and ammonia for more rapid and cost-effective analysis.
- Building qualitative models to discriminate between cultivars based on amino acid composition.

References

Bell, S-J. & Henschke, P.A., 2005. Implications of nitrogen nutrition for grapes, fermentation and wine. Australian Journal of Grape and Wine Research 11(3), 242–295.

Bely, M., Sablayrolles, J.M., Barre, P., 1990a. Automatic detection of assimilable nitrogen deficiencies during alcoholic fermentation in oenological conditions. Journal of Fermentation and Bioengineering 70(4), 246–252.

Bely, M., Sablayrolles, J.M., Barre, P., 1990b. Description of Alcoholic Fermentation Kinetics: Its Variability and Significance. Am J Enol Vitic. 41(4), 319–324.

Bernetti, I., Casini, L., Marinelli, N., 2006. Wine and globalisation: Changes in the international market structure and the position of Italy. British Food Journal 108(4), 306–315.

Bisson, L.F., 1999. Stuck and Sluggish Fermentations. Am J Enol Vitic. 50(1), 107-119.

Bisson, L.F., Waterhouse, A.L., Ebeler, S.E., Walker, M.A., Lapsley, J.T., 2002. The present and future of the international wine industry. Nature 418, 696–699.

Campbell, G. & Guibert, N., 2006. Introduction: Old World strategies against New World competition in a globalising wine industry. British Food Journal 108(4), 233–242.

Cunico, F., 2014. Innovation in the wine industry: Between change and Tradition. Thesis, Università Ca'Foscari Venezia.

Dambergs, R.G., Gishen, M., Cozzolino, D., 2015. A Review of the State of the Art, Limitations, and Perspectives of Infrared Spectroscopy for the Analysis of Wine Grapes, Must, and Grapevine Tissue. Applied Spectroscopy Reviews 50(3), 261–278.

Decanter News, 2017 Available from: https://www.decanter.com/wine-news/world-wine-production-2017-falls-oiv-378608/. Accessed September 2018.

Fleet, G., 2003. Yeast interactions and wine flavour. International Journal of Food Microbiology 86, 11–22.

Fine Wine Report 2018. Available from: World of Finance https://www.worldfinance.com/markets/fine-wine-report-2018. Accessed September 2018.

Henschke, P.A. & Jiranek, V., 1993. Yeasts - metabolism of nitrogen compounds. In Wine Microbiology and Biotechnology. Harwood Academic Publishers, 77–164.

Hui, Y.H., 2006. Handbook of Food Science, Technology, and Engineering. CRC Press.

Johnson, R. & Bruwer, J., 2007. Regional brand image and perceived wine quality: the consumer perspective. International Journal of Wine Business Research 19(4), 276–297.

Lockshin, L. & Corsi, A.M., 2012. Consumer behaviour for wine 2.0: A review since 2003 and future directions. Wine Economics and Policy 1(1), 2–23.

Pettigrew, S. & Charters, S., 2006. Product involvement and the evaluation of wine quality. Qualitative Market Research: An International Journal 9(2), 181–193.

Pretorius, I.S. & Høj, P.B., 2005. Grape and wine biotechnology: Challenges, opportunities and potential benefits. Australian Journal of Grape and Wine Research 11(2), 83–108.

Rollero, S., Bloem, A., Ortiz-Julien, A., Camarasa, C., Divol, B., 2018. Fermentation performances and aroma production of non-conventional wine yeasts are influenced by nitrogen preferences. FEMS Yeast Research 18(5), 1-11.

Swiegers, J.H., Bartowsky, E.J., Henschke P.A., Pretorius, I.S., 2005. Yeast and bacterial modulation of wine aroma and flavour Australian Journal of Grape and Wine Research 11(2), 139–173.

Torrea, D., Varela, C., Ugliano M., Ancin-Azpilicueta C., Leigh Francis I., Henschke P.A., 2011. Comparison of inorganic and organic nitrogen supplementation of grape juice – Effect on volatile composition and aroma profile of a Chardonnay wine fermented with Saccharomyces cerevisiae yeast. Food Chemistry 127(3), 1072–1083.

Vilanova, M., Ugliano, M., Varela, C., Siebert, T., Pretorius, I.S., Henschke, P.A., 2007. Assimilable nitrogen utilisation and production of volatile and non-volatile compounds in chemically defined medium by Saccharomyces cerevisiae wine yeasts. Applied Microbiology and Biotechnology 77(1), 145–157.

Chapter 2

Literature review

Unravelling the Complexities of Wine: A Big Data Approach

Chapter 2
Unravelling the Complexities of Wine: A Big Data Approach

2.1. Introduction

Wine is one of the oldest alcoholic beverages known to man and the history of winemaking has been said to parallel that of human civilization – with the earliest references dating back to 6000 BC in the Caucasus and Mesopotamia regions (Robinson & Harding, 2015). Although wine has been around for millennia, there has been a paradigm shift in the image and role of wine in society. Wine was first viewed as the only 'storable and wholesome' beverage, however, in today's times, it is perceived as a hedonistic lifestyle beverage that is often associated with an aspirational and sophisticated lifestyle (Pretorius, 2000; Bisson *et al.*, 2002; Pretorius & Bauer, 2002; Bruwer & Rueger-Muck, 2018).

Due to the growing power of the consumer in modern times (Bisson *et al.*, 2002; Deloitte Insight Report, 2014), together with the increasing awareness of food quality, safety, and authenticity (Danezis *et al.*, 2016), increasing pressure is being placed on the world-wide wine market to become more innovative to keep up with consumer demands (Pretorius & Bauer, 2002; Fleet, 2008). This is illustrated by the growing gap in supply and demand – where firstly, a global decrease in wine consumption and an increase in wine production (mainly in the New-World countries) can be observed, and secondly, the shift in consumer preferences towards more premium wines (Bisson *et al.*, 2002; Pretorius & Bauer, 2002; Swiegers *et al.*, 2005).

Thus, the statement:

> *"Wine is in the centre of the high-tension field between the forces of market pull and technology push, in which tradition and innovation need to coexist to meet the demands of wine producers and the preferences of wine consumers"*

made by Pretorius and Bauer (2002) rings true.

The '*technological push*' that is required to ensure the success of the global wine industry can take many forms, but essentially, is based on the interaction between four primary streams of knowledge and technology, namely: chemistry, biology, mechanical technologies, and scientific instrumentation (Smith, 2007). These knowledge streams are spread over both phases of the winemaking process; *i.e.* viticulture and oenology. As briefly outlined by Smith (2007), the chemistry streams include aspects such as the chemistry of the soil, the subsequent chemical reactions taking place in the vine, as well as the production and interaction of various chemical elements present in the fermenting must, and the final wine. The biological aspect entails an in-depth investigation into the biotic features such as the interactions between the various species of yeasts, bacteria, and fungi on the grape as well as during the fermentation and maturation processes. Furthermore, mechanical technologies

refer primarily to the machinery built to prune, harvest, destem, crush, and ferment grape juice to wine, while scientific instrumentation incorporates the technology required for the monitoring and control of the grape, fermenting must and wine during maturation.

Thus, at the forefront of innovation in the wine industry lies the requirement for the *deeper* understanding of the interaction of the chemical and biological constituents involved during the various stages of the winemaking process, from vine to wine. This can, in turn, be facilitated by the development of efficient, accurate and cost-effective monitoring instrumentation and protocols.

This literature review will therefore start by touching on the progression of wine research in the pursuit of quality wine production and the important role that yeast assimilable nitrogen (YAN) plays in this respect. However, due to the multitude of factors affecting the YAN concentration and composition, and the subsequent non-linear and synergistic interactions of the products of nitrogen metabolism, it is evident that, in order to allow a *holistic* understanding of the factors contributing to wine quality, a 'Big Data' approach is required. Therefore, this literature review will proceed by detailing the concept of Big Data, and what is required for this field of wine research to become part of the 'Big Data revolution'. As such, the current and prospective methods for YAN quantification are reviewed for their ability to facilitate a 'Big Data' approach to wine flavour and quality. Finally, this literature review will conclude with the need and relevance of the research conducted in the next chapters.

2.2. Wine: A conundrum

Since the discovery of the involvement of microbes in the production of wine by Louis Pasteur in 1863, scientists have become increasingly curious about the biological interactions and chemical reactions that result in the formation of this enigmatic alcoholic beverage (Pretorius, 2000).

Wine originated as a spontaneous process whereby the natural consortium of yeast present on the surface of the grape resulted in the conversion of sugars (glucose and fructose) into ethanol and carbon dioxide (Fleet, 2008). However, the understanding of this basic principle by Pasteur led to the desire of man to improve upon and control this process to their advantage, and thus, by 1890, grape juice was being inoculated with pure yeast cultures (Barnett, 2000; Pretorius, 2000; Fleet, 2008). This yeast, *Saccharomyces cerevisiae*, was selected based on its improved fermentative capacity and, subsequently, the possibility of a more predictable outcome (Swiegers *et al.*, 2005; Fleet, 2008).

As a result, a wealth of research has gone into understanding yeast and the conditions that are most conducive to the formation of a dry wine, free from spoilage. However, as this research field developed, the focus shifted towards making wines that exhibit more favourable organoleptic

qualities (Fleet, 2003; Polášková *et al.*, 2008). Due to the complexity and variability of wine and the subjectivity of human perception, extensive investigations into how consumers perceive the quality of wine have been conducted (Bisson *et al.*, 2002; Fleet, 2003). Some of the proxies that have been established as indicators of wine quality include price (Lee, 2012), awards/advice from experts (Ferro & Benito Amaro, 2018), geographical origin (McCutcheon *et al.*, 2009), absence of common wine defects/spoilage (Hopfer *et al.*, 2015), *etc*. However, at the heart of this lies the perception of the organoleptic characteristics of the wine. Thus, flavour – defined as a multisensorial construct that incorporates the sensations of the ortho- and retro-nasal olfactory systems – has been widely accepted as the *primary* proxy of wine quality (Charters & Pettigrew, 2007).

Therefore, added pressure has been placed on the wine market to produce wines that are sensorially pleasing (Swiegers *et al.*, 2005). However, the investigation of aroma in wine is not an easy task due to the varying origins and the subsequent synergistic, non-linear interactions of these sensorially active compounds (Lambrechts & Pretorius, 2000; Bisson *et al.*, 2002; Polášková *et al.*, 2008; Styger *et al.*, 2011). Wine aroma is an amalgamation of varietal aromas (from compounds originating from the grape berry), pre-fermentative aromas (due to extraction and conditioning of the grape must), fermentative aromas (produced through the metabolic activities of the yeast and bacteria) and post-fermentative aromas (that evolve during ageing of the wine due to various chemical reactions in either wooden barrels or after bottling). However, fermentative aroma compounds have been found to be the most important contributors to aroma, and, as a result, the choice of the yeast together with the fermentation conditions, are the dominant factors in determining the aroma, and subsequently, the quality of the final wine (Rapp & Versini, 1991; Lambrechts & Pretorius, 2000; Polášková *et al.*, 2008; Styger *et al.*, 2011). Therefore, as the contents of the grape berry and the resulting juice provide the nutrients required for the growth and fermentative activity of the yeast (and bacteria), the factors influencing the composition of these compounds become increasingly important in the context of quality wine production (Swiegers *et al.*, 2005).

The factors influencing the grape composition were reviewed by Jackson and Lombard (1993). These include various aspects, with varying degrees of control, many of which cannot be controlled at all – such as the macro- and meso-climate that the grapevine experiences – to factors such as micro-climate, soil and water, and competition – which can be controlled up to a point by various viticultural practices such as canopy management, irrigation and fertilization programmes, and pest, weed and disease management, respectively. Another very important factor that is reported is the genetics of both the grapevine and the rootstock which are considered to determine how the vine will react to all these aforementioned factors. Thus, the grape juice composition becomes the result of a multitude of intricate interactions, analogous to the complexity of a neural network.

2.3. YAN: A primary determinant of wine quality

One of the most important components affected by these abovementioned factors, as reviewed by Jackson and Lombard (1993), is the yeast assimilable nitrogen (YAN) concentration and composition (Bell & Henschke, 2005). As the principal yeast used for fermentation, *Saccharomyces cerevisiae,* does not exhibit sufficient extracellular proteolytic activity, it is thus not able to make use of larger peptides or grape proteins as a source of nitrogen. Thus, YAN primarily refers to α-amino nitrogen and ammonium ions, as these sources of nitrogen are able to easily pass through the yeast cell membrane (Cooper, 1982; Henschke & Jiranek, 1993; Beltran *et al.*, 2004; Bell & Henschke, 2005).

The importance of having sufficient quantities of easily assimilable nitrogen during fermentation is two-fold. Firstly, as nitrogen is required for the growth of the yeast cell by providing the necessary precursors required for protein and nucleic acid synthesis (Gobbi *et al.*, 2013), the concentration of available nitrogen significantly impacts the kinetics of the fermentation process (Bely *et al.*, 1990a; Henschke & Jiranek, 1993; Bisson, 1999). Thus, nitrogen deficiency has been highlighted as the primary cause for stuck/sluggish fermentations (Bisson, 1999). Secondly, the majority of fermentative aromas are affected by the concentration and composition of available nitrogen (Bell & Henschke, 2005; Ugliano *et al.*, 2007). The most significant impact that YAN has on wine flavour and aroma is by providing substrates (*i.e.* branched chain and aromatic amino acids) for the Ehrlich pathway (Hazelwood *et al.*, 2008). This pathway results in the formation of higher alcohols, and through subsequent reactions, various esters and volatile acids (Styger *et al.*, 2011). However, YAN has been observed to not only impact the formation of aroma compounds for which it provides direct precursors, but also in the formation of various other compounds contributing to wine flavour and aroma such as organic acids (Torrea *et al.*, 2011), and terpenes (Carrau *et al.*, 2005). In other words, YAN can be seen as central and a dominating factor in the flavour and subsequently, quality of the final wine (Ugliano *et al.*, 2007).

Thus, it is no surprise that the role of YAN in the fermentation of grape juice to wine has been an area of research that has received increasing attention in the past three decades. A bibliometric search using the *terms "Yeast Assimilable Nitrogen" AND wine OR "grape juice" OR "grape must"* as a 'topic' in the 'Web of Science database' resulted in 3113 of a total 3928 papers that were published since 1990, with more than 100 papers published annually since 2005. The increasing interest in this topic was most probably fuelled by the seminal papers published by authors such as Bely *et al.* (1990a,b) Rapp and Versini (1991), Henschke and Jiranek (1993) which, to a great extent, laid the foundations and established the importance for nitrogen research in fermentation. Specifically, the synthesis of information by Rapp and Versini (1991) reinforced the pivotal role that YAN plays in the formation of favourable flavours and aromas, and subsequently, the connection that this has on the perceived quality of the resulting wine. These ideas were echoed in a more recent review published by Ugliano *et al.* (2007).

Taking into consideration the varying origins, and the multitude of factors influencing wine flavour and aroma, it is becoming clear that wine is a multi-faceted research field on the frontier of an array of disciplines such as viticulture, microbial ecology, chemistry, and more recently, sensory science, in the pursuit of the production of a quality product, capable of meeting consumer demands (Swiegers *et al.*, 2005). In light of this, the field of wine research necessitates a collective effort to integrate all of these various streams of data in an effective and meaningful manner *i.e.* the wine research field needs to implement a Big Data approach to facilitate the deeper understanding of all the interacting factors that are at play.

The collection of a large number of samples for the purpose of understanding the nitrogen dynamics in the grapevine and the subsequent nitrogen composition of the grape juice matrix has previously been reported (Kliewer, 1970; Huang & Ough, 1991; Spayd & Andersen-Bagge, 1996; Butzke, 1998; Stines *et al.*, 2000; Nicolini *et al.*, 2004; Hagen *et al.*, 2008; Nisbet *et al.*, 2014). These investigations were carried out in the form of surveys and have either examined the nitrogen content in terms of total YAN, FAN, and ammonia (Butzke, 1998; Nicolini *et al.*, 2004; Hagen *et al.*, 2008; Nisbet *et al.*, 2014) or have taken a deeper look into the FAN content by assessing individual amino acid concentrations (Kliewer, 1970; Huang & Ough, 1991; Spayd & Andersen-Bagge, 1996; Stines *et al.*, 2000). The results were mostly presented in a descriptive format – presenting the state of the nitrogen content of different cultivars, vintages, and geographical origins in terms of average, maximum, minimum, and median values. Furthermore, the number of samples above or below a pre-determined level (of total YAN or FAN) were also reported.

The surveys on the amino acid content of grape juices also generally followed this descriptive format. However, additional investigations using the amino acid data included whether a correlation of certain amino acids (such as proline or arginine or the proline:arginine ratio), or total α-amino nitrogen or total free α-amino acids could be correlated to the amount of total soluble solids (TSS) present at harvest (Spayd & Andersen-Bagge, 1996). No correlation could, however, be found. Huang and Ough (1991) proposed that the ratio of proline:arginine can be correlated to a specific cultivar and can thus be used to discriminate between different cultivars, although this hypothesis has yet to be tested. Furthermore, Stines *et al.* (2000) studied the changes in free amino acid profiles over the course of berry ripening as well as the distribution of various amino acids between the pulp, skin and seeds at harvest. The study concluded that, due to the high arginine content of the skins, fermentation efficiency could be improved by keeping the juice in contact with skins during fermentation.

One of the major findings from the surveys was that a large percentage of samples from various cultivars and origins suffer from nitrogen deficiency (total YAN < 140 mg N/L according to (Bely *et al.*, 1990b), and are thus not capable of supporting adequate growth of yeast during fermentation. No correlation could, however, be found between FAN and ammonia concentrations and YAN was

found to be too variable to be used as an indicator of ripeness. On the other hand, Nisbet *et al.* (2014) had some success in building cultivar-specific models for the prediction of total YAN at harvest based on pre-harvest YAN levels.

Although these surveys provided value in terms of *describing* the nitrogen status, a gap still exists in the understanding of the dynamics and factors affecting/resulting in a particular YAN concentration and composition. Due to the number of compounds contributing to the YAN status, it is not surprising that the influence of the factors may be more complex to predict. Due to the highly variable and complex nature of YAN, a greater number of samples throughout the growing season may be required. However, this is only the first step. Combining a large sample set with high-throughput analytical methods and efficient statistical means of extracting information can lead to a better understanding of the evolution of this particular component of the grape juice matrix.

2.4. What is "Big Data"?

The term 'Big Data' has become a part of modern-day vocabulary, most commonly used in the field of business to facilitate the understanding of consumers. Nevertheless, the so-called 'Big-Data revolution' is just as indispensable to scientific research, providing the possibility of more data-driven and informed decision-making and hypothesis generation (Lusher *et al.*, 2014). There is, however, a rising concern among experts in this field of the understanding of what 'Big Data' really is as it is said to pave the way for the 4th industrial revolution (Yin & Kaynak, 2015). A common misconception is that size of the dataset is the only requirement that permits the use of this term (Jagadish, 2015). Therefore, many publications detailing the technicalities of Big Data have been made available (Kitchin, 2014; Lusher *et al.*, 2014; Gandomi & Haider, 2015; Jagadish, 2015; Yin & Kaynak, 2015).

The first attempt and widely accepted definition of Big Data was made by an analyst, Doug Laney from the META group. This definition came to be known as the 3Vs of Big Data: *volume, velocity* and *variety*. Later, two additional terms were added by IBM to characterise Big Data, these included *value* and *veracity* (Lusher *et al.*, 2014; Yin & Kaynak, 2015).

Volume

Simply put, *volume* refers to the magnitude of the dataset (Kitchin, 2014). However, there is a lot of debate around what constitutes as a high-volume dataset and is said to be highly dependent on the field. Thus, defining a dataset as Big Data solely on the size is widely contested (Boyd & Crawford, 2012; Jagadish, 2015). For example, data obtained in the field of social and business sciences may incorporate data from social media, video, and software programmes (Gandomi & Haider, 2015). However, datasets collected in the scientific world, especially in a field such as analytical chemistry – which requires the intentional measurement and analysis of a specific variable/compound, such

as a by-product of fermentation – are automatically orders lower. Therefore, the commonly applied criteria: that Big Data is a dataset that is 'too large to be managed by traditional methods' may not be relevant in the case of chemistry (Lusher *et al.*, 2014). This is supported by a statement made by Boyd and Crawford (2012): "*Big Data is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets*".

Velocity

Velocity is the rate at which the data is generated (Gandomi & Haider, 2015). Generally, Big Data is defined as data that is continuously being generated, enabled by technology such as smart-phones and sensors (Kitchin, 2014). Velocity is a critical aspect of Big Data as no dataset can amount to Big Data by any definition or scale if a means of obtaining the data quickly and efficiently does not exist. As chemical analyses are often complicated, time-consuming and expensive, there is a requirement of developing more easy-to-use, rapid and cost-effective means of generating data. This will facilitate the movement of analytical chemistry (and the subsequent fields for which the analysis is being conducted) to successfully enter the Big Data revolution.

Variety

Big Data is also characterised by the *variety* or heterogeneity of the type of data that is collected (Gandomi & Haider, 2015). It is by this definition that Big Data will enable the understanding of complex systems (Lehning *et al.*, 2009), such as those leading to the complex and unique flavour and quality of a wine. Obtaining a variety of data on the viticulture side is made relatively easy by infrastructure such as satellite and aerial imaging, weather stations and radars, gauge stations, ground and aerial LIDAR, temperature and moisture sensors, *etc.* provided of course, that these are correctly placed and efficiently maintained (Kitchin, 2014). In other words, systems and technology to monitor and capture the 'cause' *i.e.* the factors causing a chemical/biological change in the grape juice matrix, have already been developed. However, as eluded to above in terms of the velocity of data generation, there is a gap in the available tools to efficiently measure the 'effect' *i.e.* the chemical properties of the grape juice, fermenting must or resulting wine.

Veracity

Veracity, which refers to the reliability of the data, is a particularly major challenge of Big Data in chemistry (Lusher *et al.*, 2014; Gandomi & Haider, 2015). During chemical analysis, there is an extensive number of different factors that can lead to an erroneous measurement. Most frequently, these errors originate from sample preparation, human error, problems with equipment, calibration, reporting, calculation errors, and method selection (Ellison & Hardcastle, 2012; Committee & No 56, 2013). Therefore, data generated from different laboratories or by different operators have the possibility of being either unreliable, or if protocols are slightly modified, or a different instrument was

used, this data may not be directly comparable (Lusher *et al.*, 2014). Considering these challenges, a collaborative effort is required from the scientific community to ensure the production of high-quality data and, where ever possible, the standardisation of protocols. A movement towards this can be seen in the implementation of CODEX and VAM principles or becoming ISO accredited, which encourages the regular participation of laboratories in proficiency testing (Analytical Methods Committee Technical Brief No. 56, 2013). Thus, the adoption of these principles by laboratories conducting analysis on grape juice, must, and wine is crucial to ensure the success of Big Data in the field of viticulture and oenology. However, due to the high level of specialization that is often required for certain analyses and the logical restrictions, the standardisation of operational protocols is practically impossible.

Value

The *value* attribute of Big Data is two-fold: generally, Big Data is described as having 'low value density' *i.e.* the original form of the data has a low value in relation to its volume; however, when processed, this data can impose a great deal of value on a process or activity (Demchenko *et al.*, 2013; Gandomi & Haider, 2015). Furthermore, a paradigm shift lies between traditional data generation and analysis compared to Big Data: rather than the intentional gathering of data for a predetermined purpose, Big Data seeks to find value and gain insights from the data itself (Kitchin, 2014). Although this approach can lead to a high value impact by unveiling hidden patterns present in the data, there is a concern that, due to the magnitude of the dataset, spurious correlations can be made (Gandomi & Haider, 2015). Thus, uncorrelated variables can erroneously be found to be correlated to one another (Fan & Lv, 2008), leading to false information and, subsequently, misinformed decision-making. This is a concern that affects all fields that make use of Big data and should be acknowledged by the analyst. Fan and Lv (2008) suggest that reduction in the dimensionality of the data may help to mitigate this issue.

By reviewing the definition of Big Data, it becomes apparent that there are many dimensions in addition to just the sheer magnitude of the dataset. Furthermore, how the 5Vs of Big Data are interpreted is specific to the field. In other words, what constitutes as 'Big Data' in the field of science may not constitute as 'Big Data' in the field of business, due to the different constraints and logistics of the respective fields. For example, due to ease of data accumulation in commerce and marketing, made possible by various software packages, the internet and other digital technologies, the velocity of data accumulation far exceeds what is currently possible in certain types of scientific fields such as chemical analysis. Due to this, the volume and variety of the data is also automatically much higher. However, this emphasises the importance of high-velocity data generation in the context of Big Data. Thus, at the crux of the holistic understanding of the winemaking process by the integration of wine research into the 'Big Data revolution' lies the need for the development of *high-throughput* and *accessible* techniques for chemical analysis.

2.5. Methods currently available to measure YAN

Methods which are most commonly used for the measurement of this important component of grape juice include the Formol titration, nitrogen by *o*-phthaldialdehyde (NOPA), enzymatic ammonia, and high-performance liquid chromatography (HPLC) (Gump *et al.*, 2002). The Formol titration is a method that was first developed in 1907 by Sörensen for determining the protein concentration of samples. This method entails the addition of neutralized formaldehyde, for the purpose of liberating protons, which are subsequently titrated by sodium hydroxide to an end point, usually to a pH of 8.0. (Jodidi, 1926; Taylor, 1957; Gump *et al.*, 2002). As this method does not react with imino acids (*i.e.* secondary amino acids) such as proline and hydroxyproline, it is useful for the measurement of YAN as these amino acids are generally not assimilable by yeast under fermentative conditions (Gump *et al.*, 2002; Bell & Henschke, 2005). However, the Formol titration is only able to give an approximation of the *total* amount of YAN that is present and does not distinguish between the nitrogen contributed by amino acids (*i.e.* free amino nitrogen, FAN) and the portion of nitrogen contributed by ammonium ions (*i.e.* inorganic nitrogen).

Therefore, a method such as NOPA paired with enzymatic ammonia may be preferred as it enables the determination of not only the total amount of YAN, but the proportion of FAN to inorganic nitrogen. NOPA is able to provide a measurement of the FAN content of the must through the derivatization of α-amino acid groups with *o*-phthaldialdehyde (OPA). This results in the formation of an isoindole derivative which is quantified using a spectrometer at 335 nm (Gump *et al.*, 2002). As imino acids are not able to form the required isoindole derivative, these amino acids are also not quantified by this method. Ammonia can be spectrophotometrically quantified at 340 nm through the reaction between glutamate dehydrogenase enzyme and the ammonium ion (Dukes & Butzke, 1998).

For a more comprehensive look into the nitrogen composition of the grape juice matrix, HPLC can be used for the quantification of individual amino acids and ammonia. The use of this method for the measurement of the amino acid content of the grape juice matrix was first proposed by Dukes and Butzke (1998). This stemmed from the widespread success of the derivatization of α-amino acids with the use of OPA in various other fields of analytical chemistry. The use of OPA paired with fluorescence detection instead of the colorimetric detection of ninhydrin was an improvement first made by Roth (1971). This was done in an attempt to provide a more sensitive method, as fluorometry has been said to be "*hundred times more sensitive than colorimetry*" (Roth, 1971). In addition to the reduced sensitivity and reproducibility at low concentrations, ninhydrin is also not very robust against fluctuations in pH and temperature and is also sensitive to exposure to light and air (Callejón *et al.*, 2010). However, as previously mentioned, OPA is not able to react with imino acids, and therefore, two methods have been proposed (and are currently in use) for the quantification of these amino acids in grape juice and wine; these include the use of an additional derivatization agent, FMOCl (9H-fluoren-9-ylmethyl chloroformate) (Martínez-Rodríguez *et al.*, 2002; Beltran *et al.*, 2005;

Šuklje *et al.*, 2016), or the conversion of these (secondary) amino acids into primary amines. This is achieved by the oxidation, under alkaline conditions, of the imino acids through the addition of sodium hypochlorite (Callejón *et al.*, 2010). However, the use of AccQ•Tag as a derivatization reagent allows for the simultaneous derivatization of both ammonia and primary and secondary amino acids, and is frequently paired with ultra-performance liquid chromatography for high resolution, rapid analysis (Armenta *et al.*, 2010). The requirement for the derivatization of amino acids before detection – even though they absorb at wavelengths within the UV (190-210 nm) – is due to the interference caused by the absorption of solvents or other compounds present in the sample mixture (Callejón *et al.*, 2010). A full review of this topic can be seen in Callejón *et al.* (2010).

However, these methods are not suitable for Big Data collection. This is primarily due to the complicated protocols required for sample preparation, instrument control, and data interpretation (Liu *et al.*, 2011). Therefore, these methods can be rather labour intensive, and subsequently, time-consuming. Further disadvantages of these conventional methods include the destruction of the sample material as well as posing a threat to the environment due to use of hazardous chemicals/reagents. As a result, the generation of chemical data is slow, and usually only performed with a clear purpose or question in mind. Thus, there is a need for methods which require minimal to no sample preparation or reagents in order to provide rapid and cost-effective analysis of important components of the grape juice matrix, such as the available nitrogen (Nicolaï *et al.*, 2007; Bauer *et al.*, 2008; Gishen *et al.*, 2010; Shen *et al.*, 2010; Liu *et al.*, 2011; Cozzolino, 2015; Dambergs *et al.*, 2015).

2.6. Spectroscopy in wine research

2.6.1. Spectroscopy: A method for high-velocity data generation

The infrared (IR) region, found between the visible and microwave region of the electromagnetic spectrum, was first discovered by Herschel in 1800 (Cozzolino, 2009). The potential application of IR energy in chemical analysis was, however, only realised in 1882 by Abney and Festing, who correlated the absorption of certain wavelengths of light in this region to the presence of certain organic compounds (Thomas, 1991). Thus, an inference of the chemical composition of a particular substance/matrix can be made due to the vibrations (*i.e.* bending, stretching, rocking, scissoring, and wagging) of the chemical bonds present, and subsequently, the wavelengths of light that are absorbed versus the light that is either transmitted or reflected. This vibration of various chemical bonds at certain frequencies of IR energy is determined by properties such as the mass of the atoms, the shape of the molecule, the strength of the bonds between constituent atoms, and the periods of the associated vibrational coupling (Osborne *et al.*, 1993; Blanco & Villarroya, 2002; McClure, 2003). As a result, the need for derivatization, and possibly separation, can be eliminated, and

subsequently, a method which is both rapid and cost-effective is provided due to the minimal (or possibly no) requirement for sample preparation or reagents (Nicolaï *et al.*, 2007; Bauer *et al.*, 2008; Gishen *et al.*, 2010; Shah *et al.*, 2010; Liu *et al.*, 2011; Cozzolino, 2015; Dambergs *et al.*, 2015). This was a ground-breaking discovery, addressing all the drawbacks of conventional methods and consequently, providing a means of high-velocity data generation that is required for Big Data collection.

Furthermore, another aspect that makes spectroscopy an effective tool in the context of Big Data is the possibility of the investigation of the matrix in its *entirety* (Cozzolino *et al.*, 2009). This has several advantages above traditional methods. Firstly, together with chemometrics, the complex interactions between the various components present in the matrix can be taken into account while traditional chemical analysis tends to oversimplify the system by eliminating any interferences in the matrix (Geladi, 2003; Gishen *et al.*, 2010). This 'multivariate' approach is especially useful for a highly dynamic and complex matrix such as grapes, must, and wine (Cozzolino *et al.*, 2009). Secondly, more than one parameter can be analysed at a time, amplifying the amount of data that can be generated (Bauer *et al.*, 2008; Gishen *et al.*, 2010). Thirdly, due to the non-destructive nature of spectroscopy, *in situ* analysis of the chemical composition of the grape, must or wine is made possible, thereby enabling effective and continuous monitoring of the process (Gishen *et al.*, 2010).

2.6.2. NIR vs. MIR

The near infrared (NIR) and mid-infrared (MIR) ranges correspond to the wavenumbers 13400-4000 cm^{-1} and 4000-400 cm^{-1}, respectively (Blanco & Villarroya, 2002; McClure, 2003). The spectra obtained in the MIR range are due to the fundamental vibrations related to the stretching, bending, and rotations of chemical bonds present in the matrix. Furthermore, the MIR region can be divided into four regions corresponding to the following wavenumbers: 4000-2500 cm^{-1} (X-H stretch), 2500-200 cm^{-1} (triple bond), 2000-1500 cm^{-1} (double bond) and 1500-400 cm^{-1} (fingerprint region) (Osborne *et al.*, 1993; Blanco & Villarroya, 2002; Cozzolino, 2015). The fingerprint region is of particular interest for analysts testing the composition of various biological materials (for example, for the presence of water, proteins, lipids, fatty acids, nucleic acids and polysaccharides), as it allows for the unambiguous identification of chemical bonds. This is primarily due to the sensitivity of the bending and skeletal vibrations to large wavenumber shifts (Li-Chan, 2010).

NIR spectra on the other hand, are due to the complex overtones and combination bands of these fundamental vibrations occurring in the MIR range, and therefore, peaks in the MIR range are often much sharper, offering higher resolution than the peaks found in the NIR range (Cozzolino, 2015). Overtones occur due to anharmonic transitions between non-contiguous vibrational energy states, whereas combination bands arise from simultaneous changes in energy due to the interaction of two or more vibrational modes (Osborne, 2000; Blanco & Villarroya, 2002). The bonds most frequently observed in NIR are C-H, O-H, N-H, and S-H. This is due to the light weight of the hydrogen atom

resulting in large changes in the dipole moment, and, subsequently, large deviations from normal harmonic behaviour (Blanco & Villarroya, 2002). Consequently, the NIR spectrum is characterised by highly overlapping bands and which is said to hamper its ability to accurately measure analytes making up less than 1% of the total matrix (McClure, 2003; Cozzolino, 2015). On the other hand, overlapping spectra can lead to a reduction in the number of wavelengths that are required for the analysis of a particular compound – a potential advantage of NIR over MIR (Cozzolino, 2015) which is being facilitated by the development of increasingly advanced instruments, computers, and chemometric techniques (Kramer, 1998; Gishen *et al.*, 2010).

2.6.3. Applications of IR spectroscopy in wine research

Spectroscopy was first applied to wine in 1976 in the work done by Kaffka and Norris (Kaffka & Norris, 1976). Their work entailed the analysis in transmission mode of a small sample set of spiked red and white wines. The samples were spiked with various compounds such as ethanol, tartaric acid, and fructose. Through trial and error a set of wavelengths were eventually identified that could be used to build calibrations for the quantification of the various analytes, using multiple linear regression (MLR) analysis (Sun, 2009). However, in subsequent years, the primary use of spectroscopy in the wine industry, was for the analysis of ethanol. This has since become a standard method, used for routine analysis (Gishen *et al.*, 2010).

As spectroscopy instruments improved and the field of chemometrics developed, a range of other parameters of grapes (intact berries, berry homogenates, and juice/must) and wine (dry, sweet, dessert, and fortified) have been investigated. The progression of this field has been extensively reviewed by authors such as Gishen *et al.* (2005, 2010) Cozzolino *et al.* (2006), Bauer *et al*. (2008), Cozzolino *et al.* (2011), Dambergs *et al*. (2015), Wang *et al.* (2017), and dos Santos (2017). A great deal of emphasis has been placed on investigating the possibility of quantifying total soluble solids (TSS), total acidity (TA), pH, anthocyanins, total polyphenol content, compounds which are routinely used to determine the quality and ripeness of the berries before harvest. The rationale for the development of these calibrations is said to stem primarily from the lack of objective methods available for determining optimum harvest dates. This is a concern as the composition of the grape at harvest is accepted to be a major contributor to the quality of the final wine. Moreover, in finished wines, other than the ethanol content, the ability of spectroscopy to provide accurate readings of pH, volatile acidity, malic, tartaric, and citric acid, glycerol, reducing sugars (glucose and fructose) as well as sulphur dioxide, have also been investigated. There has also been an effort to use spectroscopy to monitor some of these parameters during fermentation. Additionally, calibrations for the concentration measurements of various trace elements such as sodium, potassium, magnesium, calcium, iron and copper have also been attempted.

In addition to quantification of important parameters present in grapes, juice/must, and wine, these reviews report how spectroscopy can be useful for qualitative analysis in wine research (Cozzolino

et al., 2006; Gishen *et al.*, 2010). The reports included an array of applications ranging from predicting wine quality scores (white and red wines) as assessed by wine experts, to the adulteration of wines from various geographical origins as well as classifying the health of grapes as well as discriminating between various yeast strains.

However, in recent years, as the field of spectroscopy and chemometrics have developed even further, the interest has shifted towards the quantification of more complex components contributing to wine quality. This is seen in the number of studies investigating the possibility of the quantification of phenolic compounds present in the skins, seeds, whole berries, berry homogenates, as well as in finished wines. The emphasis placed on these compounds stems from their fundamental role in red wine quality by contributing to the colour, mouthfeel, and flavour of these wines (Cozzolino, 2015). The degree of success of these various studies investigating the viability of different modes, wavelengths and chemometric techniques in quantifying phenolic compounds in the field of wine research has recently been reviewed by Cozzolino (2015). Further investigation into this topic has recently been conducted by Aleixandre-Tudo *et al.* (2015, 2017, 2018).

There has been much less work into the viability of using this technology as a means to quantify YAN. This is most likely due to the fact that YAN is comprised of a range of different compounds which produce a distinctly weaker signal than what can be observed for major wine compounds such as ethanol and sugar. Thus, the task of building accurate calibrations for the quantification of YAN, FAN, and ammonia is a much more daunting one. This can be seen by the unsatisfactory results reported in literature thus far. The first report for the quantification of assimilable nitrogen was by Manley *et al.* (2001). This study investigated the ability of FT-IR spectroscopy to quantify the FAN component of YAN by collecting 97 must samples from 6 different varieties over the course of two vintages. However, due to the large errors in prediction (SEP = 272.1 mg /L), rather than quantification, the study used the FAN values to discriminate between samples using Soft Independent Modelling by Class Analogy (SIMCA). Furthermore, nearly a decade later, using ATR-MIR spectroscopy, Shah *et al.* (2010) attempted to quantify total YAN as well as its components, FAN and ammonia, separately. Although this study collected a larger number of samples (n=350), these samples were only collected over a single vintage from a single winery. As such, the chances that these samples may not be representative of the variation contained by the greater population are relatively high. Furthermore, even though the errors that were obtained by Shah *et al.* (2010), were considerably lower (FAN SEP = 36.7mg N/L) than what was found by Manley *et al.* (2001), the residual prediction deviation (RPD) values that were obtained for each of the parameters were not considered adequate for the purpose of quantification (RPD = 2 for each parameter). This claim is supported by Nicolaï *et al.* (2007), who suggests that RPD values of at least 2.5 are required for a calibration to be considered acceptable for quantification purposes. Thus, Shah *et al.* (2010) also concluded a more qualitative use of their models – *i.e.* models which may be more appropriate for screening rather than quantification. Skoutelas *et al.* (2011) aimed to provide a proof of concept for

the use of FT-MIR spectroscopy for the *quantification* of total YAN. The cited study used the Formol titration as a reference method for the measurement of the total assimilable nitrogen content of 71 grape juice samples from 14 different Portuguese varieties. The partial least squares (PLS) calibrations showed very low errors in prediction (SEP = 5.9 mg N/L) and a very high RPD of 7.8. However, due to the lack of external validation and the removal of 40% (n=28 of 71) of the samples (considered by the study to be outliers), the viability of FT-IR spectroscopy for the accurate quantification of total YAN is still inconclusive.

Given the success achieved for the calibration of a complex group of compounds such as phenolics, which also have a markedly low signal in IR, as well as the central role that YAN plays in the production of quality wine, further research into this topic is warranted. However, careful consideration for the experimental design will be required. This will entail ensuring that a representative dataset is collected and that proper validation strategies are carried out to enable a realistic assessment of the predictive ability of the calibrations for the quantification of YAN, FAN, and ammonia in grape juice.

2.7. Chemometrics and calibration

In order to extract value from infrared (IR) spectroscopy, a calibration needs to be set up. This can be achieved through multivariate data analysis techniques, also known as chemometrics, which facilitates the extraction of the analytical information contained by the spectra and correlates it to the properties contained by a set of reference data (Wold, 1995; Lavine & Workman, 2006). The calibration can either be *qualitative*, allowing the grouping of samples with similar characteristics *i.e.* the *classification* of unknown samples, or *quantitative*, where the *concentration* of a particular analyte can be predicted based on the on the spectral properties (Blanco & Villarroya, 2002).

However, before chemometric techniques are applied to spectroscopic data to predict the properties of new/unknown samples, there is a number of steps that need to be taken to ensure accurate predictions can be made. Therefore, before the various multivariate techniques are discussed, these steps and the rationale behind them are briefly outlined in the following sections.

2.7.1. Gathering of calibration samples

The quality of the prediction is heavily dependent on the calibration set and, therefore, it is vitally important to ensure that the calibration set selected is representative of the population for which predictions are wished to be made (Blanco & Villarroya, 2002). The rationale for this stems from the fact that regression and classification methods used to build calibrations for spectroscopic instruments are essentially supervised learning techniques. In other words, the calibration set is the dataset that is used to *train* the model, *i.e.* the model learns from the information that is given to it in

the form of the training set, and, based on this, makes predictions on the properties of new samples that the model has not previously been exposed to (Wang *et al.*, 2012).

Due to the inherent variability of fruits and vegetables, building robust spectrophotometric calibrations for compositional analysis becomes a challenging task. Thus, the collection of a large number of samples from different 'batches' is crucial (Wang *et al.*, 1991). This means that careful consideration into what may cause variability in the sample needs to be taken into account to ensure that all this variability is well represented in the calibration set. For example, in the case of grapes for winemaking, variability may arise due to differences in cultivar and growing conditions and therefore geographical origin and vintage may also impact the variability of grape composition, in addition to the cultivar (Nicolaï *et al.*, 2007; Cozzolino, 2015; Dambergs *et al.*, 2015).

Dambergs *et al.* (2015) highlights the understanding of selecting an appropriate calibration (and validation) set as one of the largest barriers to the implementation of this technology into the wine industry. Thus, this is the first step for studies investigating the viability of this technology to provide accurate high-throughput compositional analysis for both wine research fields and the industry.

2.7.2. The use of an accurate reference method

Following the same rationale as above, whereby the calibration set is used to train the model and thus, *the quality of the prediction is based on the quality of the calibration set*, the method used to determine the reference concentrations (in the case of quantitative calibration) must be accurate (Blanco & Villarroya, 2002; Wang *et al.*, 2012). This is where the well-known phrase '*Garbage in, garbage out'* originates from in the field of chemometrics (Bakeev, 2010).

If reference methods are not carried out properly and produce values with large errors, it is possible that algorithms such as partial least squares (PLS) regression may still find correlations between these incorrect reference values and the spectra. This may lead to calibrations which seem accurate *i.e.* the reference data is faithfully represented by the model, however, in reality, the reference data does not faithfully represent the composition of the sample.

2.7.3. Recording of spectra

There are a range of considerations when deciding on which instrumentation to make use of, such as the properties of the sample to be analysed (solid/liquid/gas) as well as the appropriate wavelengths and resolution required for accurate analysis (Gishen *et al.*, 2010). A 2010 review by Gishen *et al.* stated that there were more than 65 commercially available NIR spectroscopic instruments available, which included a range of bench-top, on-line, as well as portable instruments.

The widespread application of IR spectroscopy is primarily due to the large degree of flexibility offered by these instruments, depending on the application, the characteristics of the samples, the

conditions of the surrounding environment as well as the speed of data generation that is required (Blanco & Villarroya, 2002). Broadly speaking, a IR spectrophotometer consists of a radiation source (most commonly a tungsten halogen light bulb), accessories required for sample presentation, a monochromator or interferometer, a detector, as well as a range of various optical components (optical fibres, beam splitters, integrating spheres and collimators) (Nicolaï *et al.*, 2007).

IR instruments can be grouped according to their wavelength selection properties *i.e.* whether they scan using the whole spectrum or only a limited set of fixed frequencies. Those with a limited set of frequencies either make use of filters or light emitting diodes (LEDs) and are generally simpler instruments, with limited resolution and no moving parts and are thus, generally used in portable instruments. Instruments employing the entire spectrum, generally referred to as scanning instruments, are more flexible and can therefore be used in a variety of applications. These scanning instruments can further be divided into monochromators, diode array, and Fourier-transform (FT) spectrometers. In a scanning monochromator, the individual frequencies of light are separated by either a grating or a prism (Blanco & Villarroya, 2002; McClure, 2003; Nicolaï *et al.*, 2007; Gishen *et al.*, 2010). Photodiode array (PDA) spectrometers make use of a range of diodes emitting IR radiation, and generally cover a range of 25000-5800 cm^{-1} (Osborne, 2000). There is widespread implementation of PDA spectrometers mainly due to the fast integration time and subsequent high acquisition speed, in addition to the absence of moving parts (Nicolaï *et al.*, 2007). Furthermore, FT spectrometers make use of an interferometer which modulates the radiation produced by the light source and is converted into a spectrum by means of a Fourier transform (Nicolaï *et al.*, 2007). There are two types of interferometers which are commonly used: a Michelson and a polarization interferometer, whereby the Michelson interferometer is said to produce the highest resolution (< 1 cm^{-1}) (Roberts *et al.*, 2004). Acousto-optically tunable filter (AOTF) is an additional type of monochromatic instrument, which makes use of an optical-band-pass filter that can be easily tuned to allow the passing of various wavelengths of radiation by adjusting the frequency of an acoustic wave moving through a crystal of TeO_2 (Nicolaï *et al.*, 2007). Infrared spectrometers measuring in the mid-infrared range generally make use of an interferometer (FT) and attenuated total reflection (ATR) for sample presentation (Sorak *et al.*, 2012).

In addition to the types of radiation that the sample is exposed to, the extensive number of applications of IR spectroscopy in agriculture is owed to the range of different methods available for sample presentation (Osborne, 2000). In NIR spectroscopy, this includes transmittance, reflectance, as well as hybrids of the two phenomena, transflectance and interactance (Osborne, 2000; Blanco & Villarroya, 2002; Nicolaï *et al.*, 2007). For *transmittance*, the light source is placed opposite the detector. As radiation may either be absorbed, transmitted, or reflected by the sample of interest, when the intention is to collect spectra via transmittance, reflection is eliminated and therefore, the radiation attenuated by the sample may be interpreted as transmittance. The concentration of a particular analyte of interest can then be calculated via Beer-Lambert's law. However, this law

becomes invalid in the case of light scattering, as the path length can no longer be defined due to the variation of light scattering from one sample to another. This is known as diffuse transmittance and is most commonly used for samples with a thickness of approximately 1-2 cm and are typically gathered in the range of 12500-9000 cm^{-1} (Osborne, 2000; Nicolaï *et al.*, 2007). In the case of *reflectance*, the radiation source and the detector are mounted at an angle to one another, such that the reflected radiation is recorded at an angle (for example, 45°). This is done to avoid specular reflection. Specular reflection is a phenomenon that occurs when all the radiation is reflected, and therefore, no inference can be made about the chemical composition of the sample. Diffuse reflectance on the other hand, is when scattering causes the path length to be very large, resulting in an insignificant amount of transmittance and therefore, most of the incident light rays are reflected (Osborne, 2000). *Transflectance* is a modification of this phenomenon in the case of a liquid, where a ceramic tile is placed underneath the sample. As a result, the light is transmitted through the sample, reflected by the ceramic tile, and transmitted back through the sample towards the detector. When the incident ray hits the sample surface and the resultant reflected ray is detected at a point adjacent to this incident ray, *interactance* takes place. This is achieved through the parallel placement of the light source and detector and is normally used for the analysis of large samples such as fruit (Osborne, 2000).

Attenuated total reflectance (ATR) is a technique that was developed by Fahrenfort (1961) to mitigate the issues associated with reflectance such as when substances show weak absorption but are also not suitable for transmission measurements. This was accomplished by using a dielectric with a high refractive index and the sample as the reflecting surface, and as a result, the incident ray from the highly refractive dielectric (at an angle larger than the critical angle) will be ‘totally’ reflected. This will only occur at wavelengths where the sample is non-absorbing; however, in the range where the sample is absorbing, there will no longer be total reflection, but instead, a highly contrasting and intense spectrum, similar to that of a transmission spectrum (Fahrenfort, 1961).

Furthermore, detectors in NIR spectroscopy can either be single or multiple channel. Single channel devices contain semiconductors of either PbS or InGaAs, whereas multiple channel devices contain a range of detection elements such as diode arrays (arranged in rows) or charged coupled devices (CCDs) (arranged in planes). These multi-channel devices are what facilitate the simultaneous recording of a range of wavelengths, and subsequently, responsible for the increased speed of spectra acquisition (Blanco & Villarroya, 2002).

As such, by taking all these options into account, it is clear that there are a range of factors that can affect the quality and stability of the response obtained by the spectrometer, necessitating the need for careful consideration when choosing an appropriate instrument for a specific application (Walsh *et al.*, 2000).

2.7.4. Pre-processing of spectra

The aim of pre-processing is to remove any irrelevant information or physical phenomena that may hamper the subsequent classification, multivariate regression or exploratory data analysis techniques that may be applied to the data (Rinnan *et al.*, 2009; Roussel *et al.*, 2014). However, it should be kept in mind that pre-processing is not a solution for bad data collection, but rather for the inherent issues corresponding to a specific spectroscopic technique such as the base-line shifts and non-linearities strongly associated with IR spectra (Brown *et al.*, 2000; Rinnan *et al.*, 2009; Ruah *et al.*, 2014).

Broadly, the most popular pre-processing techniques can be classified into two groups: methods for scatter-correction and spectral derivatives (Rinnan *et al.*, 2009). Scatter-correction methods are used to lessen the spectral variability between samples induced by physical phenomena and include methods such as multiplicative scatter correction (MSC), and standard normal variate (SNV). Additionally, these methods have also been observed to correct for baseline shifts. In order to remove additive and multiplicative effects, spectral derivatives can be applied. When applying the first derivative, only the baseline is removed, whereas the second derivative also removes the linear trend in addition to the baseline. However, in practice, applying derivatives to raw spectral data generally results in noise inflation. To compensate for this, the Noris-Williams and Savitzky-Golay derivation techniques were developed which optimise the signal-to-noise ratio by smoothing of the spectra (Zeaiter *et al.*, 2005; Nicolaï *et al.*, 2007; Rinnan *et al.*, 2009; Engel *et al.*, 2013).

The most effective pre-processing technique is not easy to assess before model validation. However, Rinnan *et al.* (2009) give two pieces of advice in this regard: firstly, it is not advisable to apply too many pre-processing steps to a single data set, and, secondly, essentially pre-processing should result in a *reduction* in model complexity. Furthermore, Engel *et al.* (2013) state that caution should be taken to avoid the introduction of additional variation in the data by pre-processing techniques. This statement stems from their investigation of a total of 4914 various pre-processing strategies, where only 5.6% (273) were found to reduce model complexity and subsequently, increased the model accuracy. This result reiterates the importance of proper data collection to ensure accurate predictions, rather than relying on pre-processing.

2.7.5. Chemometrics

Without the development of chemometrics, IR spectroscopy would not have been as industrially relevant as it is today. Due to the inherent multivariate nature of IR spectra, statistical techniques considering more than one variable at a time needed to be developed (Wold, 1995). Thus, in the late 1960s, extensive research was being done by an array of physical and analytical chemists to extract value from the multivariate responses obtained from these instruments, and as a result, the field of chemometrics was born (Wold, 1995; Geladi, 2003; Cozzolino *et al.*, 2009). Consequently,

chemometrics provides a means to examine as well as reveal important constituents through various interactions and interferences in the matrix (Wold, 1995; Geladi, 2003).

Chemometrics can be divided into two major categories: those used for *quantitative* analysis and those used for *qualitative* analysis (Blanco & Villarroya, 2002; Roussel *et al.*, 2014).

Quantitative Methods

Quantitative analysis is mostly used for calibration purposes making use of regression techniques *i.e.* one/more dependent variables (Y-variables) are modelled based on a set of independent response variable (X-variables). Furthermore, regression analysis is essentially an example of supervised learning as 'labelled' training data (subsequently referred to as the calibration set) is used to make an inference about future 'unlabelled' samples (Olivieri, 2018). These methods are subsequently divided into linear and non-linear methods. The most frequently used methods include multiple linear regression (MLR), principal component regression (PCR) and partial least squares (PLS) regression for linear methods and artificial neural networks (ANN) and non-linear PLS for non-linear methods (Blanco & Villarroya, 2002; Roussel *et al.*, 2014). However, this literature review will focus on briefly reviewing the linear methods.

MLR, developed by Norris in 1965, paved the way for quantitative chemometrics, however, it was not always successful at providing accurate predictions (McClure, 2003). This is mainly owed to its 'hard-modelling' approach which deals with the original variables and subsequently, assumes that the underlying chemical system is simple. In other words, the system is described in terms of a mathematical relationship whereby the measured variables are the independent variables and the outputs are the dependent variables. As a result, MLR is not robust against highly correlated (collinear), noisy data which may contain redundant X variables (Wold *et al.*, 2001; Naes *et al.*, 2002; McClure, 2003).

Due to these downfalls, soft-modelling approaches were designed by Wold, Martens and Wold (PLS) (Wold, 1975) and Cowe and McNicol (PCR) (Cowe & McNicol, 1985) which approach the regression problem from an entirely new angle. This approach assumes that the underlying chemical system is complex and therefore, soft-modelling (PLS regression and PCR) is based on the variation and correlation between the data points (*i.e.* the data found in the covariance matrix). Consequently, the interactions between variables as well as the overall variation in each of the independent variables can be taken into account (Wold, 1995; Geladi, 2003). The first step in this approach is to express the data as a set of latent variables *i.e.* the x-variables are projected onto a new set of axes which is based on the degree of variation that each x-variable contributes to, and as a result, a new set of (uncorrelated) components are derived which are orthogonal to one another. The second step in the soft-modelling approach is to eliminate the components which do not explain an adequate amount of variation in the data *i.e.* an 'optimum' number of components needs to be selected. PLS regression

is said to be superior to PCR in this regard, as the components selected in PCR are selected exclusively on the degree of predictor variance that is explained, whereas PLS regression seeks out the components that are most relevant in accurately predicting the outcome. This is important because if too many (unnecessary) components are selected, it may result in overfitting of the model, and consequently, the model will not be able to accurately predict the properties/concentration of new samples as it is too reliant on the properties of the calibration/training set. This becomes especially relevant in small datasets where the number of components selected are more than the number of available samples. In light of this, the collection of a large number of samples which represents an adequate amount of variation present in the population becomes indispensable for accurate predictions of future samples (Wold, 1995; Munck *et al.*, 1998; Osborne, 2000; Naes *et al.*, 2002; Geladi, 2003; Reiss & Ogden, 2007).

In order to ensure that the regression model will result in accurate predictions of future samples, it is imperative to validate the model (Wold *et al.*, 2001). Methods currently used for method validation include internal (cross-validation) or external (test set) validation (Consonni *et al.*, 2010). Cross-validation can be defined as a validation technique that entails the division of the dataset into a predetermined number of subsets which are iteratively left out during calibration process, which is done until all the subsets have been left out once (Hawkins *et al.*, 2003; Anderssen *et al.*, 2006). Test set validation refers to the assessment of the predictive ability of the model by an independent set of samples which were not used to develop the calibration (Golbraikh & Tropsha, 2002).

Concerns have, however, been expressed among researchers in the field of chemometrics regarding the use of cross-validation as a measure of how accurately the model will predict future samples that the model has not yet "seen" (Golbraikh & Tropsha, 2002; Anderssen *et al.*, 2006; Gramatica, 2007; Consonni *et al.*, 2010). In a compelling study done by Golbraikh and Trophsa (2002), where several published datasets were investigated, it was shown that the R^2 obtained in cross-validation (often referred to as q^2) did not correlate with R^2-values obtained using an external test set. It was found that, often, the q^2-values were over-optimistic, and when the datasets were tested with an external validation set, that the predictive ability was found to be considerably lower, yielding rather unsatisfactory results. Furthermore, Gramatica (2014) briefly overviews the arguments of experts in the field (including his own), regarding best practices for model validation. Gramatica (2014) concludes that cross-validation and test set validation should not be viewed as alternatives but rather, used sequentially. The rationale for this is that cross-validation and test set validation have completely different aims: cross-validation should be used during model *optimization* to increase the robustness of the model and to preliminarily select the best models, whereas test set validation should be used for actual *validation* of the model (Consonni *et al.*, 2010; Gramatica, 2014). Ideally, the test set should become available to the modeller after the model has been developed, however, in practice, this is often not the case due to logistical issues and additional cost. Therefore, the best chance that the modeller has to verify the predictive ability of the available model is to exploit the

data that is on hand *i.e.* splitting the dataset into a test and calibration sets (Gramatica, 2014). This test set is therefore referred to as the external validation set as these samples will not at any time be exposed to the model during optimization, but rather be used to *test* the predictive ability of the model to predict *future* samples (Gramatica, 2014). However, the problem comes in with small datasets, where, if the dataset is split, that there is a chance that the dataset that is randomly selected is predicted well due only to chance (Hawkins *et al.*, 2003; Consonni *et al.*, 2010). In these cases, Hawkins *et al.* (2003), proposes that it is more statistically sound to do cross-validation; however, cross-validation procedures should be carried out wisely.

Nevertheless, when the appropriate validation technique has been selected based on the available data and considering the logistical constraints at play, there are a few model evaluation statistics which can be used to evaluate and report on the predictive ability of the regression model. The most popular is the squared coefficient of determination, R^2 (or q^2 in the case of cross-validation). This is owed to the easy comparison between models that this parameter offers, due to the independence of this value on the scale of the specific property that is being measured (in contrast to RMSEP, for examples, which depends on the unit) (Consonni *et al.*, 2010). Instead, values universally range between 0 and 1 where 0 is indicative of the model not representing any of the variation present, whereas a value of 1 would indicate that the model accounts for the maximum amount of variation incorporated by the dataset (Consonni *et al.*, 2010). As such, more specifically, this value indicates how faithfully the variation that can be observed in the predictor variables (Y-variables) can be explained by the response variables (X-variables) in the calibration (R^2_{CAL} or q^2) and validation (R^2_{VAL} or R^2) sets (Bauer *et al.*, 2008; Aleixandre-Tudo *et al.*, 2018). Therefore, models with values closer to one are reported as having better predictive abilities and are therefore considered to be more accurate.

In addition to the squared correlation coefficient, the RMSEP (root mean square error of prediction) and RPD (residual predictive deviation) (RPD_{VAL}) can be calculated to evaluate the predictive ability of the model. This parameter is a measure of the mean deviation between the predicted and observed values (Consonni *et al.*, 2010). Thus, RMSEP is an estimate of the average uncertainty that is expected for the prediction of new samples not yet seen by the model (Nicolaï *et al.*, 2007).

The RPD is a ratio of the standard deviation incorporated by the dataset and the standard error of performance of the model, and is therefore given by the following equation:

$$RPD_{val} = \frac{Standard\ deviation_{val}}{RMSEP}$$

Consequently, the more variability incorporated in the model (*i.e.* the higher the standard deviation), and the more faithfully the model is able to predict the outcome (*i.e.* the lower the RMSEP), the higher the RPD will be, and therefore, the more reliable the model is thought to be. This is of course provided that the external validation set also incorporated enough variability to be representative of the population, allowing for realistic RMSEP values to be reported. In this case, a model with a high RPD will most likely be able to give a more accurate prediction of samples that it has not yet been exposed to. The rationale that a high standard deviation leads to more accurate prediction stems from the supervised approach of regression analysis where the model 'learns' from the characteristics presented to it in the training (calibration) set and therefore, if more information is used to train the model, the better it will be at making inferences/predictions of new samples.

Nicolaï *et al.* (2007), reviewed the RPD values that are relevant to PLS calibrations in agricultural applications. RPD values between 1.5 and 2 are thought to be only sufficient to distinguish high values from low. Although RPDs between 2 and 2.5 allow for quantification, the level of quantification is considered only rough. For acceptable quantification purposes, values above 2.5 are required and values above 3 are preferable. Shah *et al.,* (2010) regards RPD values ≥5 to be suitable for quality control for PLS calibrations for grape and wine analysis.

Qualitative methods

Other statistical methods that can be applied to chemical or spectral data are qualitative methods. These methods aim to *classify* an object (sample) rather than determining a quantitative property (Osborne, 2000). Fundamentally, these methods rely on developing a model based on pattern recognition strategies and can be divided into supervised and unsupervised techniques (Blanco & Villarroya, 2002).

Supervised methods can be divided into class-based models and discriminant analysis (DA) where class-modelling techniques focus on the *similarities* among samples in contrast to discriminant analysis which focuses on the *differences* (Blanco & Villarroya, 2002; Marini, 2010). The fundamental differences between these techniques are explained by Marini (2010) as follows: In the case of class-modelling, every class is modelled independently of the others; accordingly, each sample is either accepted or rejected by the available classes. Consequently, when there is more than one class, a particular sample may only be accepted by one of the classes; however, it is possible that the sample may be rejected by all the classes. In the case of overall rejection, this sample is identified as an outlier in terms of the available classes *i.e.* it may belong to a class that was not modelled. In contrast to this, discriminant techniques always assign a sample to one of the available classes. This is ensured by dividing the hyperspace of the available variables into as many segments as there are categories in the data. Therefore, if the coordinates of the sample fall into a particular segment which is labelled as "category 1" it will subsequently be assigned to that category. Examples of supervised methods for qualitative data analysis include Soft Independent Modelling by Class Anology (SIMCA)

supervised artificial neural networks (ANN), discriminant analysis (DA), partial-least squares discriminant analysis (PLS-DA) and its orthogonal version (OPLS-DA), and *k*-Nearest Neighbour (*k*-NN) analysis (Blanco & Villarroya, 2002; Siebert, 2011; Roussel *et al.*, 2014).

In terms of *unsupervised methods*, PCA has been acknowledged as one of the most indispensable chemometric techniques available (Cozzolino *et al.*, 2009; Siebert, 2011). The value in PCA stems from its ability to effectively screen, extract, and compress multivariate data. This is achieved through a mathematical conversion of (potentially) correlated X-variables to a set of non-correlated variables which are orthogonal to one another. As a result, the dimensionality of the data can be reduced and the components explaining the maximum amount of variance present in the dataset can be identified. Therefore, based on whether samples group together or whether they separate from one another, hidden patterns in the data can be uncovered as well as allowing the detection of outliers (Naes *et al.*, 2002; Cozzolino *et al.*, 2009; Siebert, 2011).

Cluster analysis, another important unsupervised method for qualitative chemometric analysis, can broadly be divided into hierarchical, non-hierarchical, and fuzzy clustering techniques (Siebert, 2011). The similarity between samples can be determined by various metrics including distances (Euclidean/Manhattan), correlations, as well as a combination of these. Most frequently, the samples are perceived as coordinates in a multidimensional space and the Euclidean distance between two samples are calculated; the smaller the magnitude of the distance, the more similar the samples are considered to be. The fundamental difference between hierarchical and non-hierarchical clustering is whether a relationship among the clusters is established (hierarchical) or not (non-hierarchical). Therefore, in the case of hierarchical clustering the results are often represented as a dendrogram. Hierarchical clustering can further be divided into agglomerative (bottom-up) or divisive (top-bottom) approaches whereas non-hierarchical methods can be divided into partitioning, density-based, grid-based and ‘other’ (Gülağız & Şahin, 2017). Hierarchical and non-hierarchical clustering are, however, similar in terms of the assumption of single class-membership *i.e.* each sample may belong to only one class. Conversely, fuzzy clustering algorithms allow samples to be members of two or more classes (Siebert, 2011).

The Soft Independent Modelling by Class Analogy (SIMCA) was the first supervised class-modelling method developed for the field of chemometrics (Marini, 2010). This method is in effect an extension of the unsupervised method, PCA, and is often referred to as disjoint PCA (Bauer *et al.*, 2008). This is because the method groups objects together based on applying a PCA to each class of the training set. The ideal number of PCs can be determined by either double cross-validation or amount of explained variance or in some cases, it may be pre-determined (Rácz *et al.*, 2018). Although SIMCA is a class-modelling technique, it is commonly used as a discriminatory tool in chemometrics. This is warned against by a meta-analysis conducted by Rácz *et al.* (2018), which shows that SIMCA was repeatedly outperformed for the task of discrimination by 29 different methods which includes the

majority of the major categories of the available classification methods, such as linear and quadratic discriminant analysis (LDA), Classification and Regression Tree analysis (CART), PLS-DA, *k*-NN, to name a few.

2.8. Conclusion

Due to the multi-faceted nature of the winemaking process and the increasingly competitive world wine market, a need for more innovative technologies exists. These technologies will need to enable the accurate and continuous monitoring of various aspects of the process, from vine to wine. This is important as it will provide the tools and knowledge to increase the chances that a quality product can be produced.

Due to the highly complex and variable nature of YAN, 'traditional' wine research techniques appear to be lacking in providing a comprehensive understanding of the dynamics of this important component of the grape juice matrix. A 'Big Data' approach is thus suggested as a solution to the problem. However, in order to facilitate the integration of 'Big Data' in the field of wine research, methods for more rapid and cost-effective analyses are required. In light of this, IR spectroscopy, coupled with chemometrics, is recommended as a means to measure the YAN status of the grape juice matrix. This stems from the inherent features of speed, ease-of-use, and lower costs associated with spectroscopy, in combination with the possibility of providing techniques for the multivariate assessment of complex systems, which is aided by chemometrics. Therefore, the field of chemometrics and spectroscopy, as demonstrated in the remainder of this thesis, could offer promising tools to facilitate the holistic understanding of complex systems, such as the nitrogen status of the grape juice matrix.

References

Aleixandre-Tudo, J.L., Nieuwoudt, H., Aleixandre, J.L., Du Toit, W.J., 2015. Robust Ultraviolet–Visible (UV–Vis) Partial Least-Squares (PLS) Models for Tannin Quantification in Red Wine. Journal of Agricultural and Food Chemistry 63(4), 1088–1098.

Aleixandre-Tudo, J.L., Buica, A., Nieuwoudt, H., Aleixandre, J.L., du Toit, W., 2017. Spectrophotometric Analysis of Phenolic Compounds in Grapes and Wines. J. Agric. Food Chem. 65(20), 4009–4026.

Aleixandre-Tudo, J.L., Nieuwoudt, H., Aleixandre, J.L., du Toit, W., 2018. Chemometric compositional analysis of phenolic compounds in fermenting samples and wines using different infrared spectroscopy techniques. Talanta 176, 526–536.

Analytical Methods Committee Technical Brief No. 56., 2013. What causes most errors in chemical analysis? Analytical Methods 5(12), 2914–2915.

Anderssen, E., Dyrstad, K., Westad, F., Martens, H., 2006. Reducing over-optimism in variable selection by cross-model validation. Chemometrics and Intelligent Laboratory Systems 84, 69–74.

Armenta, J.M., Cortes, D.F., Pisciotta, J.M., Shuman, J.L., Blakeslee, K., Rasoloson, D., Ogunbiyi, O., Sullivan, D.J., Shulaev, V., 2010. Sensitive and Rapid Method for Amino Acid Quantitation in Malaria Biological Samples Using AccQ•Tag Ultra Performance Liquid Chromatography-Electrospray Ionization-MS/MS with Multiple Reaction Monitoring. Analytical Chemistry 82(2), 548–558.

Bakeev, K.A., 2010. Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries. John Wiley & Sons, Chichester, UK.

Barnett, J.A., 2000. A history of research on yeasts 2: Louis Pasteur and his contemporaries, 1850-1880. Yeast 16(8), 755–771.

Bauer, R., Nieuwoudt, H., Bauer, F.F., Kossmann, J., Koch, K.R., Esbensen, K.H., 2008. FTIR Spectroscopy for Grape and Wine Analysis. Analytical Chemistry 80(5), 1371–1379.

Bell, S.-J. & Henschke, P.A., 2005. Implications of nitrogen nutrition for grapes, fermentation and wine. Australian Journal of Grape and Wine Research 11(3), 242–295.

Beltran, G., Novo, M., Rozès, N., Mas, A., Guillamón, J.M., 2004. Nitrogen catabolite repression in *Saccharomyces cerevisiae* during wine fermentations. FEMS Yeast Res 4(6), 625–632.

Beltran, G., Esteve-Zarzoso, B., Rozès, N., Mas, A., Guillamón, J.M., 2005. Influence of the Timing of Nitrogen Additions during Synthetic Grape Must Fermentations on Fermentation Kinetics and Nitrogen Consumption. Journal of Agricultural and Food Chemistry 53(4), 996–1002.

Bely, M., Sablayrolles, J.M., Barre, P., 1990a. Description of Alcoholic Fermentation Kinetics: Its Variability and Significance. Am J Enol Vitic. 41(4), 319–324.

Bely, M., Sablayrolles, J.-M., Barre, P., 1990b. Automatic detection of assimilable nitrogen deficiencies during alcoholic fermentation in oenological conditions. Journal of Fermentation and Bioengineering 70(4), 246–252.

Bisson, L.F., 1999. Stuck and Sluggish Fermentations. Am. J. Enol. Vitic. 50(1), 107-119.

Bisson, L.F., Waterhouse, A.L., Ebeler, S.E., Walker, M.A., Lapsley, J.T., 2002. The present and future of the international wine industry. Nature 418, 696–699.

Blanco, M. & Villarroya, I., 2002. NIR spectroscopy: a rapid-response analytical tool. Trends in Analytical Chemistry 21(4), 240–250.

Boyd, D., & Crawford, K., 2012. Critical Questions for Big Data. Information, Communication & Society 15(5), 662–679.

Brown, C.D., Vega-Montoto, L., Wentzell, P.D., 2000. Derivative Pre-processing and Optimal Corrections for Baseline Drift in Multivariate Calibration. Applied Spectroscopy 54(7), 1055–1068.

Bruwer, J. & Rueger-Muck, E., 2018. Wine tourism and hedonic experience: A motivation-based experiential view. Tourism and Hospitality Research 0(0), 1-15.

Butzke, C.E., 1998. Survey of Yeast Assimilable Nitrogen Status in Musts from California, Oregon, and Washington. Am. J. Enol. Vitic. 49(2), 220-224.

Callejón, R.M., Troncoso, A.M., Morales, M.L., 2010. Determination of amino acids in grape-derived products: A review. Talanta 81, 1143–1152.

Carrau, F.M., Medina, K., Boido, E., Farina, L., Gaggero, C., Dellacassa, E., Versini, G., Henschke, P.A., 2005. De novo synthesis of monoterpenes by Saccharomyces cerevisiae wine yeasts FEMS Microbiol Lett 243(1), 107–115.

Charters, S. & Pettigrew, S., 2007. The dimensions of wine quality. Food Quality and Preference 18(7), 997–1007.

Consonni, V., Ballabio, D., Todeschini, R., 2010. Evaluation of model predictive ability by external validation techniques. Journal of Chemometrics 24, 194–201.

Cooper, T.G., 1982. Nitrogen Metabolism in Saccharomyces cerevisiae. In: Strathern, J.N., Jones, E.W., Broach, J.R. (Eds.). The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp. 39-99.

Cowe, I.A. & McNicol, J.W., 1985. The Use of Principal Components in the Analysis of Near-Infrared Spectra. Applied Spectroscopy 39(2), 257–266.

Cozzolino, D., 2009. Near Infrared Spectroscopy in Natural Products. Analysis Planta Medica 75(7), 746–756.

Cozzolino, D., 2015. The Role of Visible and Infrared Spectroscopy Combined with Chemometrics to Measure Phenolic Compounds in Grape and Wine Samples. Molecules 20, 726–737.

Cozzolino, D., Dambergs, R.G., Janik, L., Cynkar, W.U., Gishen, M. 2006. Analysis of Grapes and Wine by near Infrared Spectroscopy. Journal of Near Infrared Spectroscopy 14(5), 279–289.

Cozzolino, D., Cynkar, W.U., Shah, N., Dambergs, R.G., Smith, P.A., 2009. A brief introduction to multivariate methods in grape and wine analysis. International Journal of Wine Research 1, 123-130.

Cozzolino, D., Cynkar, W., Shah, N., Smith, P., 2011. Technical solutions for analysis of grape juice, must, and wine: the role of infrared spectroscopy and chemometrics. Anal Bioanal Chem 401(5), 1475–1484.

Dambergs, R.G., Gishen, M., Cozzolino, D., 2015. A Review of the State of the Art, Limitations, and Perspectives of Infrared Spectroscopy for the Analysis of Wine Grapes, Must, and Grapevine Tissue Applied Spectroscopy Reviews 50(3), 261–278.

Danezis, G.P., Tsagkaris, A.S., Camin, F., Brusic, V., Georgiou, C.A., 2016. Food authentication: Techniques, trends & emerging approaches.Trends in Analytical Chemistry 85, 123–132.

Deloitte Insight Report, 2014. The Deloitte Consumer Review: The growing power of consumers (Consumer Review).

Demchenko, Y., Grosso, P., de Laat, C., Membrey, P., 2013. Addressing big data issues in Scientific Data Infrastructure In: International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, pp. 48-55.

Dukes, B.C. & Butzke, C.E., 1998. Rapid Determination of Primary Amino Acids in Grape Juice Using an o-Phthaldialdehyde/N-Acetyl-L-Cysteine Spectrophotometric Assay. American Journal of Enology and Viticulture 49(2), 125–134.

Ellison, S.L.R. & Hardcastle, W.A., 2012. Causes of error in analytical chemistry: results of a web-based survey of proficiency testing participants. Accreditation and Quality Assurance 17(4), 453–464.

Engel, J., Gerretzen, J., Szymańska, E., Jansen, J.J., Downey, G., Blanchet, L., Buydens, L.M.C., 2013. Breaking with trends in pre-processing? Trends in Analytical Chemistry 50, 96–106.

Fahrenfort, J., 1961. Attenuated total reflection. Spectrochimica Acta 17, 698–709.

Fan, J. & Lv, J., 2008. Sure independence screening for ultrahigh dimensional feature space. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70(5), 849–911.

Ferro, G. & Benito Amaro, I., 2018. What factors explain the price of top quality wines? International Journal of Wine Business Research 30(1), 117–134.

Fleet, G., 2003. Yeast interactions and wine flavour International. Journal of Food Microbiology 86, 11–22.

Fleet, G.H., 2008. Wine yeasts for the future. FEMS Yeast Research 8(7), 979–995.

Gandomi, A. & Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics International Journal of Information Management 35(2), 137–144.

Geladi, P., 2003. Chemometrics in spectroscopy. Part 1. Classical chemometrics. Spectrochimica Acta Part B: Atomic Spectroscopy 58, 767–782.

Gishen, M., Dambergs, R.G., Cozzolino, D., 2005. Grape and wine analysis - enhancing the power of spectroscopy with chemometrics. Australian Journal of Grape and Wine Research 11(3), 296–305.

Gishen, M., Cozzolino, D., Dambergs, R., G., 2010. The Analysis of Grapes, Wine, and Other Alcoholic Beverages by Infrared Spectroscopy. In: Handbook of Vibrational Spectroscopy, John Wiley & Sons, Ltd. Chichester, UK. pp 1-18.

Gobbi, M., Comitini, F., D'Ignazi, G., Ciani, M., 2013. Effects of nutrient supplementation on fermentation kinetics, H2S evolution, and aroma profile in Verdicchio DOC wine production. European Food Research and Technology 236(1), 145–154.

Golbraikh, A. & Tropsha, A., 2002. Beware of q2! Journal of Molecular Graphics and Modelling 20, 269–276.

Gramatica, P., 2007. Principles of QSAR model validation: internal and external. QSAR & Combinatorial Science 26(5), 694–701.

Gramatica, P., 2014. External Evaluation of QSAR Models, in Addition to Cross-Validation: Verification of Predictive Capability on Totally New Chemicals. Molecular Informatics 33(4), 311–314.

Gülağız, F.K. & Şahin, S., 2017. Comparison of Hierarchical and Non-Hierarchical Clustering Algorithms. International Journal of Computer Engineering and Information Technology 9(1), 6-14.

Gump, B.H., Zoecklein, B.W., Fugelsang, K.C., Whiton, R.S., 2002. Comparison of Analytical Methods for Prediction of Pre-fermentation Nutritional Status of Grape Juice. Am. J. Enol. Vitic. 53(4), 325–329.

Hagen, K.M., Keller, M., Edwards, C.G., 2008. Survey of Biotin, Pantothenic acid, and Assimilable Nitrogen in Winegrapes from the Pacific Northwest. Am. J. Enol. Vitic. 59(4), 432–436.

Hawkins, D.M., Basak, S.C., Mills, D., 2003. Assessing Model Fit by Cross-Validation. Journal of Chemical Information and Computer Sciences 43(2), 579–586.

Hazelwood, L.A., Daran, J.-M., Maris, A.J.A. van, Pronk, J.T., Dickinson, J.R., 2008. The Ehrlich Pathway for Fusel Alcohol Production: A Century of Research on Saccharomyces cerevisiae Metabolism. Appl. Environ. Microbiol. 74(8), 2259–2266.

Henschke, P.A. & Jiranek, V., 1993. Yeasts - metabolism of nitrogen compounds. In: Wine Microbiology and Biotechnology. Harwood Academic Publishers, Chur, Switzerland. pp. 77–164.

Hopfer, H., Nelson, J., Ebeler, S.E., Heymann, H., 2015. Correlating Wine Quality Indicators to Chemical and Sensory Measurements. Molecules 20(5), 8453–8483.

Huang, Z. & Ough, C.S., 1991. Amino Acid Profiles of Commercial Grape Juices and Wines. Am. J. Enol. Vitic. 42(3), 261–267.

Jackson, D.I. & Lombard, P.B., 1993. Environmental and Management Practices Affecting Grape Composition and Wine Quality - A Review. Am. J. Enol. Vitic. 44(4), 409–430.

Jagadish, H.V., 2015. Big Data and Science: Myths and Reality Big Data Research 2(2), 49–52.

Jodidi, S.L., 1926. The Formol Titration of Certain Amino Acids. J. Am. Chem. Soc. 48(3), 751–753.

Kaffka, K.J., Norris, K.H., 1976. Rapid Instrumental Analysis of Composition of Wine. Acta Alimenta 5, 267-279.

Kitchin, R., 2014. Big Data, new epistemologies and paradigm shifts. Big Data & Society April-June 1-12.

Kliewer, W.M., 1970. Free Amino Acids and Other Nitrogenous Fractions in Wine Grapes. Journal of Food Science 35(1), 17–21.

Kramer, R., 1998. Chemometric techniques for quantitative analysis. Marcel Dekker, Inc, New York.

Lambrechts, M.G. & Pretorius, I.S., 2000. Yeast and its Importance to Wine Aroma - A Review S. Afr. J. Enol. Vitic. 21(Special Issue), 97-129.

Lavine, B. & Workman, J., 2006. Chemometrics. Analytical Chemistry 78(12), 4137–4145.

Lee, F.S., 2012. Wine and the Consumer Price-perceived Quality Heuristics. International Journal of Marketing Studies 4(3), 31-35.

Lehning, M., Dawes, N., et al., 2009. Instrumenting the earth: next generation sensor networks in environmental science. In: In T. Hey, S. Tansley, & K. Tolle (Eds.), The Fourth Paradigm: Data-Intensive Scientific Discovery. pp. 45–51.

Li-Chan, E.C.Y., 2010. Introduction to Vibrational Spectroscopy in Food Science In: J.M. Chalmers & P.R. Griffiths (Eds.), Handbook of Vibrational Spectroscopy. John Wiley & Sons, Ltd, Chichester, UK.

Liu, F., He, Y., et al., 2011. Detection of Organic Acids and pH of Fruit Vinegars Using Near-Infrared Spectroscopy and Multivariate Calibration. Food and Bioprocess Technology 4(8), 1331–1340.

Lusher, S.J., McGuire, R., van Schaik, R.C., Nicholson, C.D., de Vlieg, J., 2014. Data-driven medicinal chemistry in the era of big data. Drug Discovery Today 19(7), 859–868.

Manley, M., van Zyl, A., Wolf, E.E.H., 2001. The Evaluation of the Applicability of Fourier Transform Near-Infrared (FT-NIR) Spectroscopy in the Measurement of Analytical Parameters in Must and Wine. South African Journal of Enology & Viticulture 22(2), 93-100.

Marini, F., 2010. Classification Methods in Chemometrics. Current Analytical Chemistry 6(1), 72–79.

Martínez-Rodríguez, A.J., Carrascosa, A.V., A.V., Martín-Álvarez, P.J., Moreno-Arribas, V., Polo, M.C., 2002. Influence of the yeast strain on the changes of the amino acids, peptides and proteins during sparkling wine production by the traditional method. Journal of Industrial Microbiology & Biotechnology 29(6), 314–322.

McClure, W.F., 2003. 204 Years of near Infrared Technology: 1800–2003. Journal of Near Infrared Spectroscopy 11(6), 487–518.

McCutcheon, E., Bruwer, J., Li, E., 2009. Region of origin and its importance among choice factors in the wine-buying decision making of consumers. Intl Jnl of Wine Business Res 21(3), 212–234.

Munck, L., Nørgaard, L., Engelsen, S.B., Bro, R., Andersson, C.A., 1998. Chemometrics in food science—a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. Chemometrics and Intelligent Laboratory Systems 44, 31–60.

Naes, T., Isaksson, T., Fearn, T., Davies, T., 2002. A User-Friendly guide to Multivariate Calibration and Classification. NIR Publications, Chichester UK.

Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007. Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology 46(2), 99–118.

Nicolini, G., Larcher, R., Versini, G., 2004. Status of yeast assimilable nitrogen in Italian grape musts and effects of variety, ripening and vintage. Vitis 43(2), 89–96.

Nisbet, M.A., Martinson, T.E., Mansfield, A.K., 2014. Accumulation and Prediction of Yeast Assimilable Nitrogen in New York Wine-grape Cultivars. American Journal of Enology and Viticulture 65(3), 325–332.

Olivieri, A.C., 2018. Introduction to Multivariate Calibration: A Practical Approach. Springer International Publishing.

Osborne, B.G., 2000. Near-Infrared Spectroscopy in Food Analysis In: R.A. Meyers (Eds.), Encyclopedia of Analytical Chemistry. John Wiley & Sons, Ltd, Chichester, UK.

Osborne, B.G., Fearn, T., Hindle, P.H., 1993. Practical NIR spectroscopy with applications in food and beverage analysis. (2nd Ed). Addison-Wesley Longman Ltd: Harlow UK.

Polášková, P., Herszage, Ebeler, S.E., 2008. Wine flavor: chemistry in a glass. Chem. Soc. Rev. 37(1), 2478–2489.

Pretorius, I.S., 2000. Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. Yeast 16(8), 675–729.

Pretorius, I.S. & Bauer, F.F., 2002. Meeting the consumer challenge through genetically customized wine-yeast strains Trends in Biotechnology 20(10), 426–432.

Rácz, A., Gere, A., Bajusz, D., Héberger, K., 2018. Is soft independent modeling of class analogies a reasonable choice for supervised pattern recognition? RSC Advances 8(1), 10–21.

Rapp, A. & Versini, G., 1991. Influence of nitrogen compounds in grapes on aroma compounds of wines In: Developments in Food Science 37. Elsevier 1659–1694.

Reiss, P.T. & Ogden, R.T., 2007. Functional Principal Component Regression and Functional Partial Least Squares Journal of the American Statistical Association 102(479), 984-996.

Rinnan, Å., van Den Berg, F., Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. Trends in Analytical Chemistry 28(10), 1201–1222.

Roberts, C.A., Workman, J., Reeves, J.B., 2004. Near-infrared spectroscopy in agriculture. In: Al-Amoodi, L., Roberts, C.A., Reeves III, J.B., (eds) Number 44 in the Series: Agronomy. American Society of Agronomy, Inc., Crop Science Society of America Inc., Soil Science Society of America, Inc. Publishers Madison, Wisconsin, USA.

Robinson, J. & Harding, J., 2015. The Oxford Companion to Wine. Oxford University Press.

Roth, M., 1971. Fluorescence reaction for amino acids. Anal. Chem. 43(7), 880–882.

Roussel, S., Preys, S., Chauchard, F., Lallemand, J., 2014. Multivariate Data Analysis (Chemometrics) In: C.P. O'Donnell, C. Fagan, & P.J. Cullen (eds). Process Analytical Technology for the Food Industry. Springer New York, New York, NY 7–59.

Ruah, M.E.N.M., Rasaruddin, N.F., Fong, S.S., Jaafar, M.Z., 2014. Data pre-processing methods of FT-NIR spectral data for the classification cooking oil. AIP Conference Proceedings 1653(1) 890–897.

dos Santos, C.A.T., Páscoa, R.N.M.J., Lopes, J.A., 2017. A review on the application of vibrational spectroscopy in the wine industry: From soil to bottle. Trends in Analytical Chemistry 88, 100–118.

Shah, N., Cynkar, W., Smith, P., Cozzolino, D., 2010. Use of Attenuated Total Reflectance Midinfrared for Rapid and Real-Time Analysis of Compositional Parameters in Commercial White Grape Juice. Journal of Agricultural and Food Chemistry 58(6), 3279–3283.

Shen, F., Niu, X., Yang, D., Ying, Y., Li, B., Zhu, G., Wu, J., 2010. Determination of Amino Acids in Chinese Rice Wine by Fourier Transform Near-Infrared Spectroscopy. J. Agric. Food Chem. 58(17), 9809–9816.

Siebert, K.J., 2011. Using Chemometrics To Classify Samples and Detect Misrepresentation In: S.E. Ebeler, G.R. Takeoka, & P. Winterhalter (eds). Progress in Authentication of Food and Wine. Vol. 1081. American Chemical Society, Washington, DC 39–65.

Skoutelas, D., Ricardo-da-Silva, J.M., Laureano, O., 2011. Validation and Comparison of Formol and FT-IR Methods for Assimilable Nitrogen in Vine Grapes. South African Journal of Enology and Viticulture 32(2), 262–266.

Smith, K., 2007. Technological and economic dynamics of the world wine industry: An introduction. International Journal of Technology and Globalisation 3(2/3), 127-137.

Sorak, D., Herberholz, L., Iwascek, S., Altinpinar, S., Pfeifer, F., Siesler, H.W., 2012. New Developments and Applications of Handheld Raman, Mid-Infrared, and Near-Infrared Spectrometers. Applied Spectroscopy Reviews 47(2), 83–115.

Spayd, S.E. & Andersen-Bagge, J., 1996. Free Amino Acid Composition of Grape Juice From 12 Vitis vinifera Cultivars in Washington. Am. J. Enol. Vitic. 47(4), 389–402.

Stines, A.P., Grubb, J., Gockowiak, H., Henschke, P.A., HøJ, P.B., Heeswijck, R., 2000. Proline and arginine accumulation in developing berries of Vitis vinifera L. in Australian vineyards: Influence of vine cultivar, berry maturity and tissue type. Australian Journal of Grape and Wine Research 6(2), 150–158.

Styger, G., Prior, B., Bauer, F.F., 2011. Wine flavor and aroma. J. Ind. Microbiol. Biotechnol. 38(9), 1145–1159.

Šuklje, K., Antalick, G., Buica, A., Langlois, J., Coetzee, Z.A., Gouot, J., Schmidtke, L.M., Deloire, A., 2016. Clonal differences and impact of defoliation on Sauvignon blanc (Vitis vinifera L.) wines: a chemical and sensory investigation: Sauvignon blanc wine: chemical and sensory investigation. Journal of the Science of Food and Agriculture 96(3), 915–926.

Sun, D.-W., 2009. Infrared Spectroscopy for Food Quality Analysis and Control. Academic Press.

Swiegers, J.H., Bartowsky, E.J., Henschke, P.A., Pretorius, I.S., 2005. Yeast and bacterial modulation of wine aroma and flavour. Australian Journal of Grape and Wine Research 11(2), 139–173.

Taylor, W.H., 1957. Formol Titration: An Evaluation of its Various Modifications. Analyst 82(976), 488–498.

Thomas, N.C., 1991. The early history of spectroscopy Journal of Chemical Education 68(8), 631.

Torrea, D., Varela, C., Ugliano M., Ancin-Azpilicueta C., Leigh Francis I., Henschke P.A., 2011. Comparison of inorganic and organic nitrogen supplementation of grape juice – Effect on volatile composition and aroma profile of a Chardonnay wine fermented with Saccharomyces cerevisiae yeast. Food Chemistry 127(3), 1072–1083.

Ugliano, M., Henschke, P.A., Herderich, M.J., Pretorius, I.S., 2007. Nitrogen management is critical for wine flavour and style. Wine Industry Journal. 22(6), 24-30.

Walsh, K.B., Guthrie, J.A., Burney, J.W., 2000. Application of commercially available, low-cost, miniaturised NIR spectrometers to the assessment of the sugar content of intact fruit. Australian Journal of Plant Physioogy 27, 1175–1186.

Wang, L., Sun, D.-W., Pu, H., Cheng, J.-H., 2017. Quality analysis, classification, and authentication of liquid foods by near-infrared spectroscopy: A review of recent research developments. Critical Reviews in Food Science and Nutrition 57(7), 1524–1538.

Wang, S., Wu, D., Liu, K., 2012. Semi-supervised Machine Learning Algorithm in Near Infrared Spectral Calibration: A Case Study to Determine Cetane Number and Total Aromatics of Diesel Fuels In: IEEE 308–311.

Wang, Y., Veltkamp, D.J., Kowalski, B.R., 1991. Multivariate instrument standardization. Analytical Chemistry 63(23), 2750–2756.

Wold, H., 1975. Soft Modelling by Latent Variables: The Non-Linear Iterative Partial Least Squares (NIPALS) Approach perspectives in Probability and Statistics: Papers in Honour of M. S. Bartlett. 117–142.

Wold, S., 1995. Chemometrics; what do we mean with it, and what do we want from it? Chemometrics and Intelligent Laboratory Systems 30(1), 109–115.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58(2), 109–130.

Yin, S. & Kaynak, O., 2015. Big Data for Modern Industry: Challenges and Trends [Point of View] Proceedings of the IEEE 103(2), 143–146.

Zeaiter, M., Roger, J.-M., Bellon-Maurel, V., 2005. Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods. Trends in Analytical Chemistry 24(5), 437–445.

Chapter 3

Research Results

A Statistical Exploration of Survey Data to Identify the Role of Cultivar and Origin in the Concentration and Composition of Yeast Assimilable Nitrogen

Chapter 3
A Statistical Exploration of Survey Data to Identify the Role of Cultivar and Origin in the Concentration and Composition of Yeast Assimilable Nitrogen

3.1. Introduction

Nitrogen is a crucial soil-derived macronutrient that plays a central role in the metabolic processes of both grapevine and yeast. Yeast are, however, only able to make use of a certain portion of the nitrogen contained by the grape berry and this is commonly referred to as Yeast Assimilable Nitrogen (YAN). These assimilable nitrogen sources include ammonium and free alpha-amino nitrogen (FAN). Proteins and high molecular weight peptides are not classified as a source of YAN as yeast do not possess the sufficient extracellular proteolytic activity that is required to catabolise these nitrogen sources into an assimilable state. Furthermore, the lack of oxygen present during fermentation prevents the breakdown of secondary amino acids such as proline and hydroxyproline and thus, these amino acids are not capable of serving as a source of YAN (Bell & Henschke, 2005).

The concentration and composition of YAN has frequently been reported to influence the quality of the grape, and consequently, of the final product (Filipe-Ribeiro & Mendes-Faia, 2007; Mendes-ferreira, 2011) This is in part due to the critical role that YAN plays in the kinetics of fermentation. YAN has been identified as the primary determinant of fermentation rate when all other nutrients are supplied in sufficient quantities (Bell & Henschke, 2005; Beltran *et al.*, 2005; Gobbi *et al.*, 2013). As YAN provides the amino acids that are required for protein synthesis of the yeast cell, sufficient YAN will ensure adequate biomass production and a successful fermentation (Hernandez-Orte *et al.*, 2006). Therefore, low YAN concentrations have frequently been identified as the cause for stuck and sluggish fermentations (Bisson, 1999). A 140 mg N/L level has been identified as the minimum concentration of YAN required to ferment to dryness a clarified must of moderate sugar concentration, and thus, a grape juice containing less than 140 mg N/L can be classified as nitrogen deficient (Bely *et al.*, 1990). Furthermore, low YAN concentrations have also been linked to the production of reductive aromas as an increase in H_2S is produced in the absence of adequate S-containing amino acids such as cysteine and methionine (Henschke & Jiranek, 1993; Barbosa *et al.*, 2012).

Consequently, nitrogen supplementation in the form of diammonium phosphate (DAP) and complex nutrients containing variable concentrations and compositions of amino acids has become a common practice in wineries. These additions are often made without prior knowledge of the concentration and composition of YAN available for fermentation. Prophylactic additions primarily

stem from the lack of access of many wineries to skilled personnel and equipment required to analyse this grape must parameter, or due to the logistical issues with time as wineries may not receive the results in a suitable time-frame to allow for appropriate supplementation decisions (Gump *et al.*, 2002). Moreover, the consequences of having very high YAN concentrations – such as microbial instability, the production of carcinogens (ethyl carbamate), allergens (biogenic amines) and haze-causing proteins – are rarely considered by industry (Bell & Henschke, 2005). This is most probably due to winemakers being more regularly confronted with the symptoms that are associated with nitrogen deficiency.

Due to the essential role that YAN plays during fermentation, it is important to understand the nature of YAN as well as the factors affecting its assimilation and metabolism within the grapevine. YAN has been identified as a highly variable component of grape juice, and said to be affected by a variety of factors such as cultivar, climate, vintage, soil, and various viticultural processes such as the trellising system employed, soil and canopy management as well as rate, timing and form of nitrogen application (Bell & Henschke, 2005).

Surveys of YAN of *Vitis vinifera* species have been conducted in various winemaking-regions of the world. The first was conducted on the 1996 vintage on the West Coast of the United states (Butzke, 1998). This study, however, only gave an overview of the YAN concentrations found for seven different varieties. In 2004, Nicolini *et al.* investigated the variability of YAN with respect to some of the factors said to influence its concentration and composition. These factors included cultivar, vintage, and different growing regions located in the Trentino region of North East Italy. YAN, in addition to biotin and pantothenic acid concentrations was also surveyed in the Pacific Northwest and significant differences were reported between cultivars, vintage, and vineyard locations (Hagen *et al.*, 2008). Nisbet *et al.* (2014), conducted a three-year study in New York state for the purpose of building cultivar-specific regression models to predict harvest YAN based on pre-harvest YAN measurements.

Furthermore, the prediction of YAN concentration and composition through the use of infrared (IR) spectroscopy, coupled with chemometrics, has become an increasingly attractive concept due to its rapid and cost-effective nature (Gishen *et al.*, 2005; Shah *et al.*, 2010). This task, however, requires a representative data set for robust calibration of the IR instrument (Patz *et al.*, 2004).

Currently, no large-scale studies of the nitrogen status (YAN concentration and composition) of grape juices exists for the South African wine industry. Moreover, due to the variable nature of YAN, the information gathered in these surveys may not necessarily be applicable in a South African context. Therefore, it would be beneficial to gain insight into nitrogen status of the South African wine industry.

However, even though cultivar, vintage, and growing region were included as parameters in determining the variance of YAN concentrations in the previous surveys, their relative importance in determining the YAN concentration and composition has not yet been investigated. Thus, the primary aim of this study was to gain insight into the YAN status of the South African wine industry, however, due to the number of samples that were subsequently collected, an investigation into the role that cultivar and growing district, as demarcated by the Wine of Origin System in South Africa, was conducted.

This chapter therefore explores the statistical methods that would be appropriate to elucidate the roles of the aforementioned variables in determining the concentration and composition of YAN. Furthermore, this study aims to inform the research community of the nature of this important component of grape juice – aiding future studies involving YAN.

3.2. Materials and Methods

3.2.1. Sample Collection

A survey of the YAN status of 805 commercial grape juices was conducted over the 2016 and 2017 harvest. This survey followed an unsupervised format and therefore, no specific grape cultivar or grape-growing district was targeted for the purpose of this study. However, due to logistical issues, grape-growing districts outside of the Western Cape region of South Africa were not considered. Furthermore, the origin of the grapes was determined according to grape-growing district as demarcated by the Wine of Origin System (SAWIS, 2018).

Settled grape juice samples were collected from wineries after crushing from grapes harvested at a ripeness level suitable for commercial winemaking. Upon collection, samples were coded and stored at -20°C until analysis. Of the 363 samples collected in 2016, 343 could be identified according to cultivar and 318 according to the district of origin. In 2017, all samples (n=442) could be identified according to the origin and 395 according to cultivar.

3.2.2. Analytical Methods

The components of YAN, free amino nitrogen (FAN) and ammonia were measured separately by enzymatic assay using the Megazyme™ K-PANOPA (Ireland) for FAN and Enzytec™ Fluid Ammonia (Id-No: E5390, R-Biopharm, Germany) for Ammonia. This was performed on the Arena 20XT (Thermo Fisher Scientific, Waltham, MA) which provides automated spectrophotometric readings. These individual values for FAN and ammonia were then summed to determine the total amount of YAN available and expressed as mg N/L (Dukes & Butzke, 1998).

3.2.3. Statistical Analysis

Variance tests were carried out in the statistical software SPPS (IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp). Clustering analysis was conducted using RStudio version 1.1.442 (RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA http://www.rstudio.com/). Principal component analysis (PCA) as well as classification and regression tree (CART) analysis were performed using the statistical software package STATISTICA (version 13, TIBCO Software Inc. 2017, http://statistica.io).

3.3. Results and Discussion

3.3.1 YAN: A comparative study

YAN concentrations were found to be highly variable (11-fold), ranging from 45 to 484 mg N/L. Variation was also observed for the individual components of YAN: FAN and ammonia. FAN concentrations varied between 30 and 365 mg N/L while ammonia varied between 0 and 167 mg N/L (Table 3.1). This highly variable nature of YAN has been expressed by previous studies such as those conducted on the West Coast of North America (14-fold) (Butzke, 1998) and in the Trentino region of North East Italy (25-fold) (Nicolini *et al.*, 2004). The variation found in the present study was not only observed between cultivars but also within cultivars, with most cultivars having standard deviations of 22-30% of the mean. This may be owed to the different ripening status of samples of the same cultivar. This variation is in corroboration with findings of Nicolini *et al.* (2004), who found YAN to be the most variable component of grape juice that is used to assess the quality of the grape at harvest. Overall, the average YAN concentration was found to be 191 ± 64 mg N/L, with FAN averaging 138 ± 46 mg N/L and ammonia 53 ± 24 mg N/L (Table 3.1). Pinotage had the highest average YAN of 348 ± 77 mg N/L (Figure 3.1, Table A3.1). This is approximately 100 mg N/L more than the average for the highest white YAN-yielding cultivar, Viognier (250 ± 56 mg N/L), and two- to three-fold more average YAN than for most other red cultivars (Cabernet Franc and Merlot: 3-fold and Cabernet Sauvignon and Shiraz: 2-fold) investigated in this study (Figure 3.1; Table A3.1). Pinotage is a cultivar that was created in South Africa in 1925 by crossing two species of *Vitis vinifera*, Pinot Noir and Cinsaut (http://pinotage.co.za). As only four samples of Pinot Noir were collected in this survey, it was not used to make any conclusions; however, all samples of Pinot Noir collected had YAN values between 194 and 283 mg N/L (data not shown). Furthermore, Pinot Noir has consistently been reported to have high average YAN levels by other surveys, conducted in different wine regions of the world (Butzke, 1998; Nicolini *et al.*, 2004; Nisbet *et al.*, 2014). To our knowledge, no other data has been published concerning YAN levels of Cinsaut grapes, however, in this study, it was identified as the red grape cultivar to have the second highest average YAN of

194 ± 63 mg/L (n=15) (Figure 3.1, Table A3.1). Due to the parent varietals of Pinotage having higher average YAN levels than other red cultivars surveyed, it may suggest a genetic basis for YAN and thus, YAN may be a cultivar-specific trait. Several authors have eluded to cultivar playing a role in the level of YAN that can be found in the grape juice before the start of fermentation (Christensen, 1984; Huang & Ough, 1989), albeit affected by other factors such as rootstock, soil, climate, vintage and viticultural practices (Bell & Henschke, 2005).

Table 3.1. Descriptive Statistics of YAN, FAN, and ammonia concentrations for all commercial grape juice samples (n=805) collected over the course of the survey.

	Average	***Median***	***Min***	***Max***	***Lower Quartile***	***Upper Quartile***
YAN	191±64	184	45	484	145	233
FAN	138±46	131	30	365	106	166
Ammonia	53±24	52	0	167	34	68

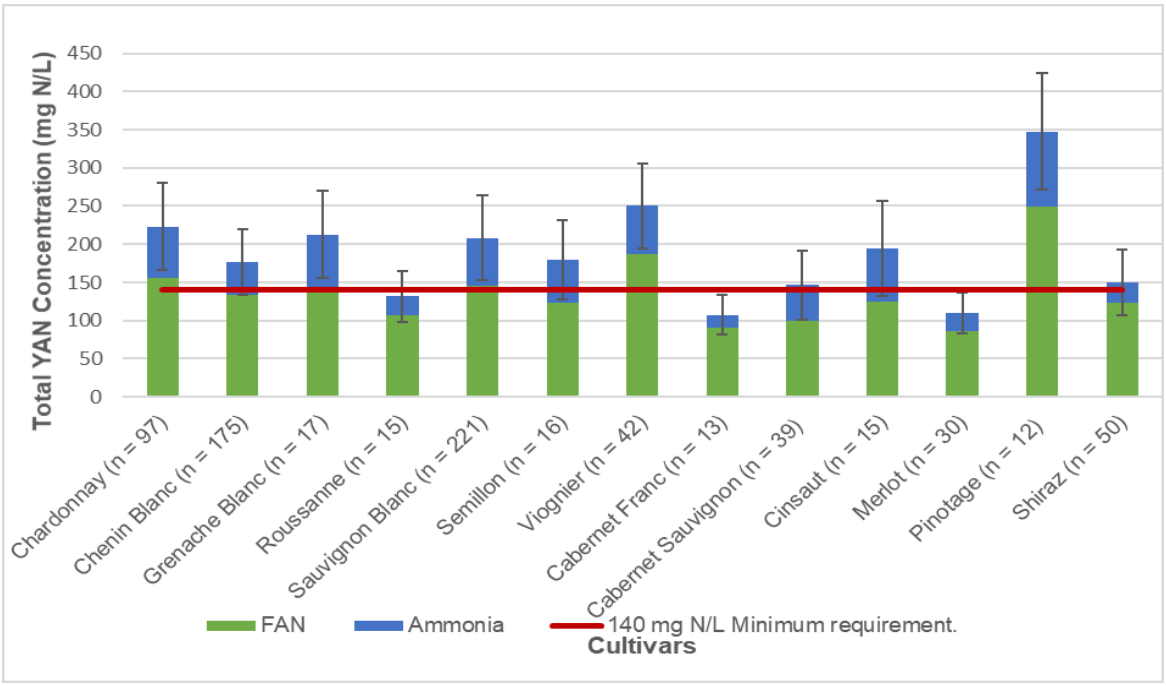


Figure 3.1. Average YAN concentrations represented as a stacked plot of FAN and ammonia concentrations for cultivars contributing more than 10 samples over the course of the 2016 and 2017 harvest.

The impact that cultivar may have on resulting YAN levels is further illustrated by similar trends that could be observed in other YAN surveys, irrespective of where in the world the survey was conducted. Other than Pinot Noir, Chardonnay was also highlighted to have consistently high levels of YAN, on most occasions averaging more than 200 mg N/L (Butzke, 1998; Nisbet *et al.*, 2014), which is on par with the average found for South African Chardonnay grape juices (223 ± 57 mg N/L) (Figure 3.1, Table A3.1). The same trend for YAN levels in Chardonnay grape juices has been reported in Trentino, Italy, for five different vintages between 1986 and 1996 (Nicolini *et al.*, 2004).

Hagen *et al.* (2008), who found an average YAN less than 200 mg N/L for Chardonnay grape juices surveyed in the Pacific Northwest (165 mg N/L), still reported it to have the highest average YAN compared to the other cultivars included in the survey.

On the other hand, Cabernet Franc exhibited the lowest average YAN (108 ± 26 mg N/L), with Merlot following closely (110 ± 26 mg N/L) (Figure 3.1, Table A3.1). Subsequently, these cultivars had more than 90% of their juices containing less than the recommended level of 140 mg N/L of YAN (Figure A3.1). Cabernet Franc has regularly been identified as the cultivar which has the lowest average YAN concentration, regardless of the cultivars it was compared to (Butzke, 1998; Nisbet *et al.*, 2014b). Furthermore, Nisbet *et al.* (2014), found the average YAN concentration for this cultivar to be as low as 75 ± 44 mg N/L. As such, it seems that Cabernet Franc may be a cultivar that frequently suffers from nitrogen deficiency. This should be taken note of by winemakers as this cultivar will most likely require nitrogen additions to help prevent the occurrence of stuck fermentations. The same may be true for Merlot as four different vineyards that were surveyed over three vintages in the Pacific Northwest repeatedly found juices with average YAN levels less than 140 mg N/L (Hagen *et al.*, 2008). On several occasions, the average was even found to be below 100 mg N/L (Hagen *et al.*, 2008). Although higher than the average of 110 ± 26 mg N/L found for Merlot in this study (Figure 3.1, Table A3.1), YAN levels for Merlot in New York State were also reported to be on average deficient (132 ± 47 mg N/L) (Nisbet *et al.*, 2014).

Other cultivars which seem to suffer from nitrogen deficiency include Roussanne and Cabernet Sauvignon, having average YAN concentrations of 132 ± 34 mg N/L and 146 ± 45 mg N/L, respectively (Figure 3.1; Figure A3.1). Overall, 23% of South African grape juices were found to be deficient. This is less than half than what was observed in the survey conducted in North East Italy, where approximately 58% of the 586 juices surveyed were found to have YAN levels below the threshold level of 140 mg N/L (Nicolini *et al.*, 2004). On the contrary, only 13% of 1523 grape juices surveyed on West Coast of the U.S. were found to be nitrogen deficient (Butzke, 1998). The large discrepancy between the studies is most likely due to the (i) varying number of different cultivars that were included in the different surveys, (ii) the large variation of climate, soil, rootstocks and viticultural practices said to play a role in determining the concentration of YAN (Bell & Henschke, 2005) as well as (iii) the variable nature of this component of grape juice, as established by the various YAN surveys (Butzke, 1998; Nicolini *et al.*, 2004; Hagen *et al.*, 2008; Nisbet *et al.*, 2014).

On the whole, even though the absolute YAN values were found to be different between the various surveys, what is striking is the similarity in ranking. When taking into account the cultivars found to be common between the present study and various other surveys, YAN concentrations were always found to follow the following rank: Chardonnay > Merlot > Cabernet Franc (Butzke, 1998; Nisbet *et al.*, 2014). Interestingly, the ranking of Chardonnay > Syrah > Cabernet Sauvignon > Merlot found by Hagen *et al.* (2008), in the Pacific Northwest was found to be exactly the same in this study of

South African grape juices. Although Butzke (1998), found Cabernet Sauvignon (192 ± 32 mg N/L) concentrations to be lower than Merlot (196 ± 36 mg N/L), Sauvignon Blanc was still shown to have a higher average YAN than both of these red cultivars as well as having an average YAN lower than Chardonnay – once again corroborating the results of the present study.

A correspondence analysis (CA) was used to identify which cultivars associated with specific levels of YAN, namely ‘very low’, ‘low’, ‘high’ and ‘very high’ (Figure 3.2). These levels were subsequently designated by conducting a cluster analysis on the cases. The first dimension accounted for 72.40% of the total inertia and dimension 2 for 25.59%. Dimension 1 was primarily characterised by the positive loadings of Viognier which are associated with ‘very high’ levels of YAN and the negative loadings of Cabernet Franc, Merlot, Cabernet Sauvignon and Roussanne, associated with ‘very low’ levels of YAN. Dimension 2 was mainly driven by the strong positive loadings of Pinotage, also associated with ‘very high’ levels of YAN, whereas negative loadings on dimension 2 were characterised by Chenin Blanc, Semillon and Cinsaut which associate with ‘low’ levels of YAN. Chardonnay, Grenache Blanc and Sauvignon Blanc, grouping together with ‘high’ levels of YAN, had slightly positive loadings along dimension 1. However, there does not seem to be a large distinction between groups of ‘high’ and ‘low’. Furthermore, the biggest variation seems to lie between the red and white cultivars investigated in this survey. As such, the correspondence analysis was able to give a concise overview of the structure of the observations and variables in the dataset, substantiating the trends found thus far.

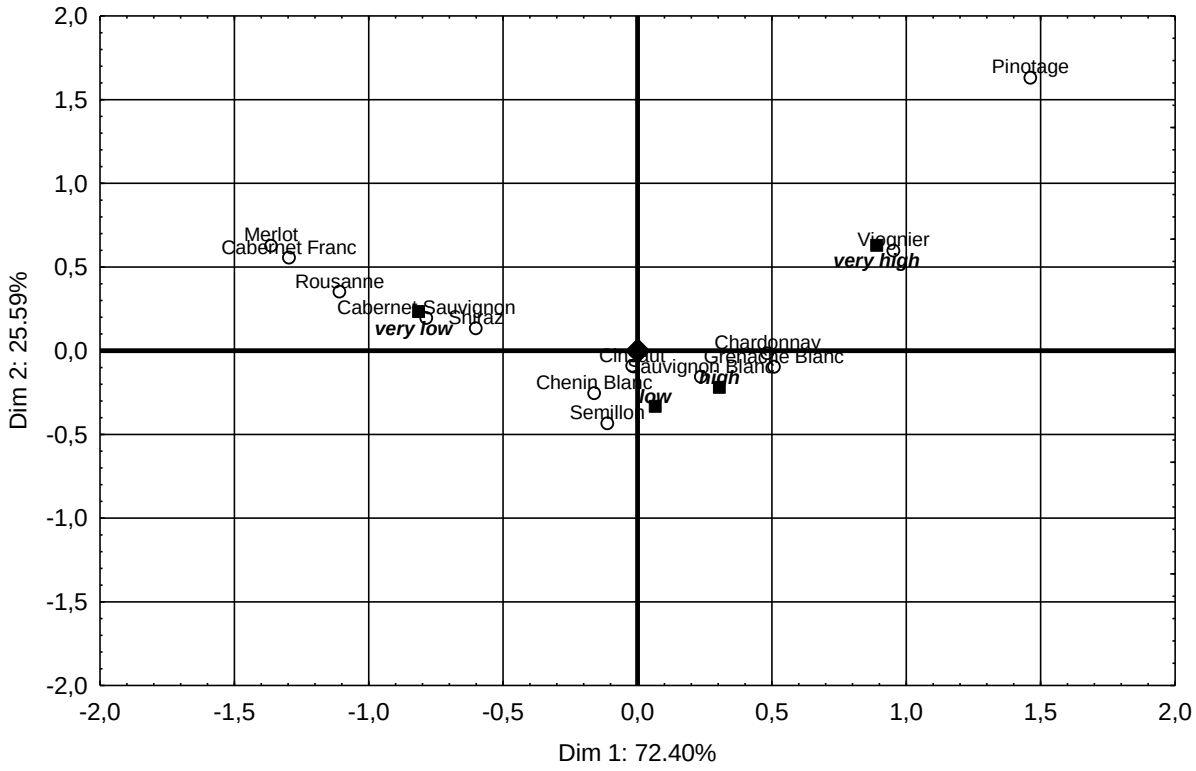


Figure 3.2. Correspondence analysis of cultivars contributing more than 10 samples over the two vintages associating with 'very low', 'low', 'high' and 'very high' levels of YAN.

3.3.2. Role of cultivar

Whether cultivar plays a role in the resulting YAN concentration and composition can be investigated through an analysis of variance between cultivars. If significant differences are found between the mean concentrations of YAN, FAN and ammonia, it could be said that cultivar is a driving force in determining the concentration and composition of YAN, irrespective of the extrinsic factors that the grapevine is exposed to.

However, because of the unsupervised nature of this study, the data set presents some difficulties to answer this question. ANOVA (one-way analysis of variance) is a test which is readily used by scientists to test the equality of means. However, the robustness of this test can be severely affected when there is a violation of certain assumptions either separately or in conjunction with one another (Harwell *et al.*, 1992; Coombs *et al.*, 1996). These assumptions include independence, normality, and homogeneity of variance (Moder, 2007). Although ANOVA has been found to be reasonably tolerant of data that is not perfectly normally distributed (Lindman, 1992), it has been shown to perform poorly when the homogeneity of variance assumption is not adhered to. This is because ANOVA on heteroscedastic data has been observed to increase the chances of obtaining a type I error (Box & Anderson, 1955).

A type I error occurs when the null hypothesis is rejected when there are actually no statistically significant differences between the groups tested (Betz & Gabriel, 2011). In this case, it would mean that cultivars were found to be statistically significantly different from one another based on their YAN, FAN, and ammonia concentrations when, in reality, they are not. As such, cultivar would falsely be identified as a role-player in determining the final YAN concentration and composition. Furthermore, the danger of incorrectly finding significant differences between groups, and committing a type I error, becomes further exacerbated when comparing groups of unequal sizes (Ahad & Yahaya, 2014).

Therefore, the first step when attempting to do an ANOVA would be to establish whether the dataset is appropriate for this type of analysis by testing whether the necessary assumptions have been adhered to. Normality was tested by performing a Shapiro-Wilk Test, whose null hypothesis states that the data is normally distributed (Shapiro & Wilk, 1965). Thus, $p<0.05$ would indicate that the dataset is not normally distributed. The values of the Shapiro-Wilk Test for YAN, FAN, and ammonia concentrations were all found to be highly significant and as such, the dataset does not follow a normal distribution for any of the tested variables (Table 3.2). Furthermore, this dataset was also found to violate the assumption of homogeneity of variance as Levene's null hypothesis (Levene, 1960) of equal variances across groups was rejected by obtaining a p-value < 0.05 (Table 3.2). As such, the dataset can be described as non-parametric and heteroscedastic and, consequently, a traditional one-way ANOVA would not be appropriate to test for statistically significant differences

between cultivars. This is in addition to the unequal sample sizes that enhances the probability of obtaining a type I error. Therefore, an alternative test needed to be applied to this dataset to increase the chances of honestly determining the role of cultivar in determining the YAN concentration and composition.

Table 3.2. *p*-Values for the respective tests for cultivars contributing more than 30 samples to the survey. Values indicated in red are significant at $p < 0.001$

	Shapiro-Wilk	***Levene's Test***	***Welch's t-Test***
YAN	0.000	0.000	0.000
FAN	0.000	0.000	0.000
Ammonia	0.000	0.000	0.000

Welch's *t*-Test appeared to be an ideal option as it does not assume equal variances as the denominator in the equation is not based on a pooled variance estimate but rather allows for heterogeneity of the variance to be taken into account (Ahad & Yahaya, 2014). Thus, the Welch's *t*-Test protects against a type I error in the case of heteroscedasticity. However Algina *et al.* (2011), found that Welch's *t*-Test may have difficulty controlling for type I error when the dataset was found to be non-normal. This was further investigated by Ahad & Yahaya (2014), who tested the sensitivity of the Welch's *t*-Test in different scenarios of heteroscedasticity and normality. This study found that Welch's *t*-Test was reasonably robust and capable of protecting against type I error rates in the case of non-normal data when variance and group sizes were positively paired. As such, it should be noted that when data is not normally distributed, whether the heterogeneity of variance between groups is positively or negatively paired with group sizes is another factor that was shown to affect the robustness of the Welch's *t*-Test. As a *positive* pairing of variance and sample size was observed in this survey (Table A3.2) the use of the Welch's *t*-Test was deemed appropriate to reduce the chance of a type I error in our data, and honestly report whether there are significant differences between cultivars. However, as a precautionary measure, only cultivars which contributed more than 30 samples were included for Welch's *t*-Test. This was decided as it will allow the Central Limit Theorem to come into effect. (The Central Limit Theorem works on the premise that irrespective of the distribution, that when the sample size is larger than 30, that the distribution of the sample *means* will eventually result in a normal distribution (Hildebrand, 2008)).

Therefore, the white cultivars Chardonnay, Chenin Blanc, Sauvignon Blanc, and Viognier and the red cultivars Cabernet Sauvignon, Merlot, and Shiraz were considered for this analysis. Furthermore, Welch's *t*-Test has shown increased robustness against unequal sample sizes compared to one-way ANOVA (Ahad & Yahaya, 2014). The outcome of the Welch's *t*-Test showed that cultivars were indeed significantly different from one another ($p<0.05$) (Table 3.2). Therefore, we fail to reject the

hypothesis that cultivar is the major role-player in determining the YAN concentration and composition irrespective of the extrinsic factors that the grapevine it is exposed to.

In order to elucidate which cultivars were responsible for the significant differences, a post-hoc test was performed. Due to the highly variable nature of YAN, FAN, and ammonia concentrations and the resulting non-normal distribution of the data, it was decided to use the Games-Howell nonparametric test. Moreover, Games-Howell shows reasonable robustness against unequal sample sizes paired with heteroscedastic data and, as with the Welch's *t*-Test, protects against a type I error especially in the case of positive pairing of sample sizes and variance. Furthermore, Games-Howell has exhibited superior power in comparison to the Tukey post-hoc test as it was observed to provide better protection against Type II errors in the aforementioned case (Rusticus & Lovato, 2014). A type II error occurs when the null hypothesis is retained when in reality it is false *i.e.* finding no significant differences when in fact there are. Therefore, the lower the power of the test, the higher the chances are of obtaining a type II error (Lieberman & Cunningham, 2009). Thus, conducting a Games-Howell Test gave the highest probability of obtaining a 'truthful' result with this particular dataset. However, it must be kept in mind that these tests are never able to accept a hypothesis, but rather, gives the possibility to "fail to reject" the hypothesis.

The Games-Howell post-hoc test showed that on average, YAN, FAN, and ammonia concentrations for different cultivars were found to be significantly different from one another, however, a few exceptions were observed (Table A3.3). These exceptions could be seen for Cabernet Sauvignon and Shiraz, Chardonnay and Sauvignon Blanc, as well as Chardonnay and Viognier for total YAN. In terms of FAN, Chardonnay and Sauvignon Blanc, Chenin Blanc and Shiraz, and Cabernet Sauvignon and Merlot were not found to be significantly different. Ammonia, however, showed more similarities between cultivars. Ammonia concentrations were not found to be different between Chenin Blanc and Cabernet Sauvignon, Merlot and Shiraz, and Viognier and Chardonnay as well as between Viognier and Sauvignon Blanc. More significant differences were most probably found for YAN and FAN due to the complex nature of FAN – as FAN is the total of amino acids present in grape juice (with the exception of proline and hydroxyproline) and YAN being the sum of FAN and ammonia (Bell & Henschke, 2005).

Before moving forward, one should reflect on what these statistical results may mean from a biological perspective, Myles *et al.* (2011), who assessed the genetic diversity and population structure of 583 species of *Vitis vinifera,* found that the majority of the cultivars that they investigated were found to be part of a single pedigree. Furthermore, it was found that the domestication of wine grapes involved an extremely weak bottleneck and therefore, there was a minimal reduction in genetic diversity. As such, many cultivars may share a large number of traits. Despite this, most grapevine cultivars considered in this study showed significant differences between their concentration and composition of YAN, and thus, it is hypothesised that processes related to nitrogen

assimilation and metabolism of the grapevine are in most cases cultivar-specific and not a universally shared trait between all species of *Vitis vinifera*.

3.3.3. Role of origin

In 1974, The Wine of Origin System demarcated different regions, districts and wards within South Africa based on different land types (Southey, 2017). Land types are considered to be different when there is a difference between one or more of the following factors: macroclimate, terrain form or soil pattern. Districts are demarcated around geographic structures such as rivers and mountain ranges (Saayman, 1999).

To determine whether the origin of the grapes *i.e.* the district (as demarcated by the Wine of Origin System) where the grapes were grown played a role in the level of YAN, FAN, or ammonia, a Welch's *t*-Test was again conducted. For this test, only cultivars that contributed 30 or more samples were considered. Furthermore, only districts that contributed 5 or more samples of a particular cultivar were considered for this analysis. Therefore, to make use of a parametric test such as Welch's *t*-Test with less than 30 samples per district being compared, the normality of the data first needed to be established. Only if *p*-values for the Shapiro-Wilk Test were found not to be significant ($p>0.05$) and the dataset was indeed found to be normally distributed, were further analysis conducted to test the effect of district on the YAN, FAN, or ammonia levels of a particular cultivar. Data sets not found to be normally distributed were not included in the Welch's *t*-Test. Moreover, all datasets were found to adhere to the assumption of homogeneity of variance except for all variables tested for Sauvignon Blanc and for the ammonia concentrations for Chenin Blanc (Table 3.3). Therefore, the Welch's *t*-Test was employed instead of ANOVA to detect any statistically significant differences between different districts for a particular cultivar. Welch's *t*-Test resulted in $p<0.05$ for all the variables for all the white cultivars investigated (Table 3.3). YAN, FAN, and ammonia concentrations for all the red cultivars (Cabernet Sauvignon, Merlot and Shiraz), with the exception of FAN for Shiraz ($p = 0.030$) were not found to be statistically significantly different between any of the districts considered in this study (Table 3.3).

Table 3.3. *p*-Values for the respective tests for districts contributing more than 30 samples and cultivars contributing >5 samples per district over the course of the 2016 and 2017 harvest. Values indicated in red are significant at $p < 0.05$.

Test	*Chardonnay*	*Chenin Blanc*	*Sauvignon Blanc*	*Viognier*	*Cabernet Sauvignon*	*Merlot*	*Shiraz*
Shapiro-Wilk							
YAN	0.518	0.886	0.058	0.253	0.798	0.169	0.431
FAN	0.317	0.668	0.002	0.714	0.202	0.058	0.499
Ammonia	0.577	0.217	0.411	0.006	0.017	0.487	0.007
Levene's Test							
YAN	0.101	0.057	0.005	0.189	0.133	0.816	0.548
FAN	0.336	0.271	0.015	0.131	0.105	0.371	0.409
Ammonia	0.159	0.003	0.000	0.162	0.263	0.899	0.489
Welch's t-Test							
YAN	0.000	0.009	0.000	0.002	0.868	0.218	0.090
FAN	0.001	0.007	0.000	0.004	0.495	0.164	0.030
Ammonia	0.002	0.001	0.002	0.007	0.948	0.045	0.402

The reasons for more significant differences for the white cultivars than for red were further investigated by having a closer look at the raw data. The red cultivars included in this analysis (Cabernet Sauvignon, Merlot, and Shiraz) were obtained from four districts, namely: Franschhoek, Paarl, Stellenbosch, and Swartland (*i.e.* these districts all contributed >5 samples per cultivar and were thus included in the analysis) (Tables A3.4, A3.5, A3.6) Geographically, these districts are all in close proximity to one another.

When looking at the white cultivars (Chardonnay, Chenin Blanc, Sauvignon Blanc and Viognier), it was found that, generally, the significant differences for YAN concentrations were between 'cooler' and 'warmer' climate districts (Tables A3.7, A3.8, A3.9, respectively). For example, Sauvignon Blanc, a very well represented cultivar in this survey (n=221), with 198 samples included in this specific analysis, showed significant differences between districts such as: Elgin (cooler) and Darling, Paarl and Swartland (warmer), Overberg (cooler) and Swartland (warmer) (Table A3.9). Furthermore, Chardonnay showed highly significant differences between Elgin (cooler) and two warmer climate districts, namely Paarl and Stellenbosch (Table A3.7). Conversely, the only significant differences found for Chenin Blanc were between two 'warmer' climate districts (Paarl and Stellenbosch) (Table A3.8). However, the *p*-value obtained between Paarl and Stellenbosch for the YAN levels of Chenin Blanc was found to be rather close to the α-value of 0.05 ($p = 0.045$).

How far a district is from the ocean plays a major role in the districts' climate (Myburgh, 2005). This is mainly due to the circulation of winds between the land and the sea and the phenomenon of continentality. As such, districts closer to the ocean will have cooler climates which are more resistant to temperature increases and fluctuations due to the increased effective heat capacity of water bodies compared to dry land (Bonnardot *et al.*, 2002; Southey, 2017). Therefore, cooler climate

areas, such as Elgin and Overberg are denoted as such as they are mostly found closer to the ocean, whereas Franschhoek, Paarl, Swartland, and Stellenbosch are classified as warmer climate districts due to their increased distance from the ocean. These theories are substantiated by the increased annual mean maximum temperatures found for inland districts such as Franschhoek compared to Elgin which is found closer to the coast (Vink *et al.*, 2010). Moreover, temperature has been highlighted as one of the most important variables in viticulture, having major effects on all the physiological processes of the grapevine and subsequently, the final grape juice composition (Carey, 2001; Myburgh, 2005). Therefore, less statistically significant differences may have been found for red cultivars as they were all contributed by warmer (inland) districts.

However, as the current study entailed an unsupervised survey of such a large number of commercial grape juices, the specific climatic factors of each particular vineyard were not recorded. Furthermore, obtaining accurate climatic data of the various winemaking districts is made especially difficult by not only the sparseness of the weather stations across the Western Cape of South Africa, but also due the poorly chosen locations of these existing weather stations (Southey, 2017). Moreover, Southey (2017), often found the data obtained from these various weather stations to be unreliable. Therefore, the role of climatic factors on the concentration and composition of YAN of a particular cultivar could not be elucidated in the current study.

Nevertheless, based on these results, the particular districts as demarcated by the Wine of Origin System do not seem to play a statistically significant role in the composition and concentration of YAN of a particular cultivar. Rather, the proximity of the district to the ocean and, subsequently, the climate associated with the district is hypothesised to have an effect. However, the role of climate in the total concentration and composition of YAN can only be investigated with the support of appropriate climatic data.

Relatedness between cultivars

Clustering is an unsupervised method which groups objects together based on their similarity/dissimilarity through the simultaneous consideration of a given set of variables (Bailey, 1975). Due to the unsupervised nature of this technique, it is able to reveal ‘hidden patterns’ which are able to represent a ‘data concept’, providing a concise summary of the data at hand (Berkhin, 2006). Thus, performing a clustering analysis on the present data would enable the determination of the similarity/dissimilarity of the cultivars based on their YAN, FAN, and ammonia concentrations. These results can then be compared to the genetic relationships of various species of *Vitis vinifera* published in the seminal paper by Myles *et al.* (2011).

Before starting a cluster analysis, a few decisions needed to be made to decide on the method that would best suit the data and purpose of the analysis. These included whether each object may belong to only one cluster (exclusive) or whether it could belong to more than one (nonexclusive). If

an exclusive technique is decided upon, the researcher needs to decide whether a sequential or simultaneous formation of clusters would be most appropriate. In sequential formation of clusters, hierarchical and non-hierarchical methods exist. Hierarchical techniques enable the relationships between clusters to be defined whereas non-hierarchical methods just aim to achieve a minimum within-cluster variance. Both techniques can follow an agglomerative or divisive method of clustering. Generally, agglomerative methods follow a polythetic strategy which can be defined as a strategy where each object may possess a large number of characteristics which may be shared by a large number of individuals within the cluster; however, no one characteristic is necessarily possessed by every member of the cluster. On the contrary, divisive methods may largely follow a monothetic strategy where all characteristics *need* to be possessed by a member in order for it to be included in the cluster (Bailey, 1975, 1983).

The decision-making process on which clustering analysis to use does not always need to follow a top-to-bottom approach. Although, if a decision is made further down, the 'above' features will automatically become features of the analysis. For example, as the purpose is to establish the relationships and thus, the relatedness of cultivars, a hierarchical clustering analysis was decided upon and, subsequently, an exclusive and sequential clustering strategy is employed. Furthermore, the dataset (of 13 cultivars) is small enough for easy and meaningful interpretation of hierarchical clustering analysis.

Whether to use agglomerative or divisive methods is not subsequent to the decision to use hierarchical clustering and therefore, would be the next decision that needs to be made. According to Sokal & Sneath (1963), agglomerative methods are generally better at obtaining natural groups in comparison to divisive methods due to their inherent polythetic nature. This rationale stems from their extensive research in taxonomic classification where they argue that natural systems are, in most cases, polythetic and sometimes even fully polythetic – where no one characteristic is shared by all members of a cluster. Therefore, using a divisive strategy is not considered to be appropriate for natural systems as it is thought to be largely monothetic. Monothetic strategies can be especially problematic when an object is classified as being distant based on only a single characteristic, where they would otherwise be classified as being similar based on all the other characteristics in consideration (Sokal & Sneath, 1963).

Due to the vastly variable nature of YAN and to the fact that no correlation between FAN and ammonia ratios could be found (Butzke, 1998) (and therefore, FAN and ammonia are considered to be polythetic characteristics), it was decided to take an agglomerative, rather than a divisive approach. Agglomerative clustering uses a "bottom-up" approach where, at the start, each object constitutes its own cluster and new clusters are formed by successively merging the most similar clusters into bigger clusters (Bailey, 1975). This is done by iteratively calculating a similarity-dissimilarity matrix to establish the relationship between the new cluster and the remaining objects

in the property space (Blashfield, 1976). The most widely used algorithms for hierarchical clustering analysis can be broken down into two groups: (i) linkage methods and (ii) methods where the cluster centres can be specified (either as an average or a weighted average of the objects within the cluster). Linkage methods include single, complete, and average, whereas the latter includes methods such as centroid, median, and minimum variance (also known as Ward's method) (Murtagh & Legendre, 2011).

The methods used to form the clusters can, however, yield vastly different results and thus, any solution obtained should be tested. How accurately a dendrogram represents the pairwise distances between the original (unmodeled) data can be evaluated by the cophenetic correlation coefficient (CCC) (Sokal & Rohlf, 1962). This coefficient can be defined as a correlation between the Euclidean distance/dissimilarity between a pair of observations and their corresponding cophenetic distance/dissimilarity. The cophenetic distance is the intergroup distance which 'allowed' the observations to be merged together into one cluster (Saraçli *et al.*, 2013).

Three different summary statistics were used to represent the data – the mean, the median, and the interquartile range. The median and the interquartile range were also considered for the clustering analysis as it is more robust against the presence of outliers and nonparametric datasets.

The dendrograms that were constructed were chosen based on the results of the CCC calculations (Table A3.10). For the current dataset, the average (for the interquartile range and the mean) and centroid (for the median) linkage methods performed the best, yielding the highest CCC values. Similar trends were observed by (Saraçli *et al.*, 2013) in a simulation study evaluating the best clustering methods for an array of distance measures. Therefore, the relationships represented in these dendrograms were further investigated.

The median and mean using centroid and average linkage methods, respectively, yielded the same relationships among clusters and thus, in this case, the difference between the median and the mean as well as average and centroid linkage methods appeared to be negligible in terms of the clustering analysis. The average linkage method can be defined as the average between all the pairwise distances between the two clusters while the centroid method is the distance between the means of each cluster. The interquartile range however, yielded a slightly different dendrogram. Therefore, in effect, two different dendrograms could be used to investigate the relationships among cultivars based on their YAN, FAN, and ammonia concentrations (Figure 3.3 and 3.4).

These dendrograms could now be used as a basis to investigate whether the relationships exhibited between cultivars based on their YAN concentration and composition was similar to their genetic relationships. Granted, any conclusions made from this comparison will rely on inductive reasoning rather than deductive reasoning as no gene expression studies were conducted during the survey.

Myles *et al.* (2011), found closer genetic relationships between cultivars such as Chardonnay and Pinot Noir, both associated with high YAN levels, than between Chardonnay and cultivars such as Cabernet Sauvignon, Cabernet Franc and Merlot, which have all been observed to be associated with lower levels of YAN (Butzke, 1998; Nicolini *et al.*, 2004; Hagen *et al.*, 2008; Nisbet *et al.*, 2014).

The most striking similarity between the current study and Myles *et al.* (2011), was the association of Merlot and Cabernet Franc. In both dendrograms (Figure 3.3 and 3.4) and in the study conducted by Myles *et al.* (2011), these two cultivars were found to be very closely related. Genetically, Merlot was found to share a first degree (parent-offspring) relationship with Cabernet Franc (Myles *et al.*, 2011). Furthermore, the sibling cultivars Shiraz and Viognier grouped together in a cluster using the interquartile range with average linkage. Moreover, Cinsaut and Pinotage, who share a parent-offspring relationship, were found together in cluster G in the cluster analysis based on the interquartile range (Figure 3.3). However, Pinotage was found to be in a cluster of its own, away from all of the other cultivars, in the dendrogram based on the mean and the median (Cluster A; Figure 3.4).

On the whole, the cluster analysis using the interquartile range with average linkage showed stronger parallels with the study done by Myles *et al.* (2011). Due to the differences in cultivars that were included in the two studies, it is difficult to make a complete direct comparison; however, similar trends could be seen for the cultivars considered in both studies.

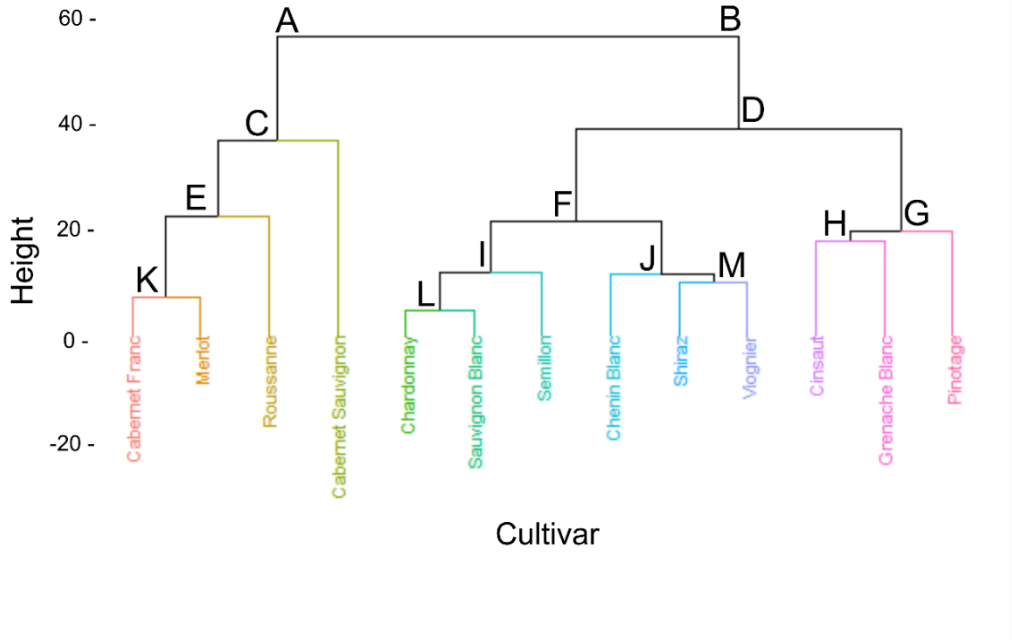


Figure 3.3. Dendrogram of the hierarchical agglomerative clustering analysis using the interquartile range and the average linkage method

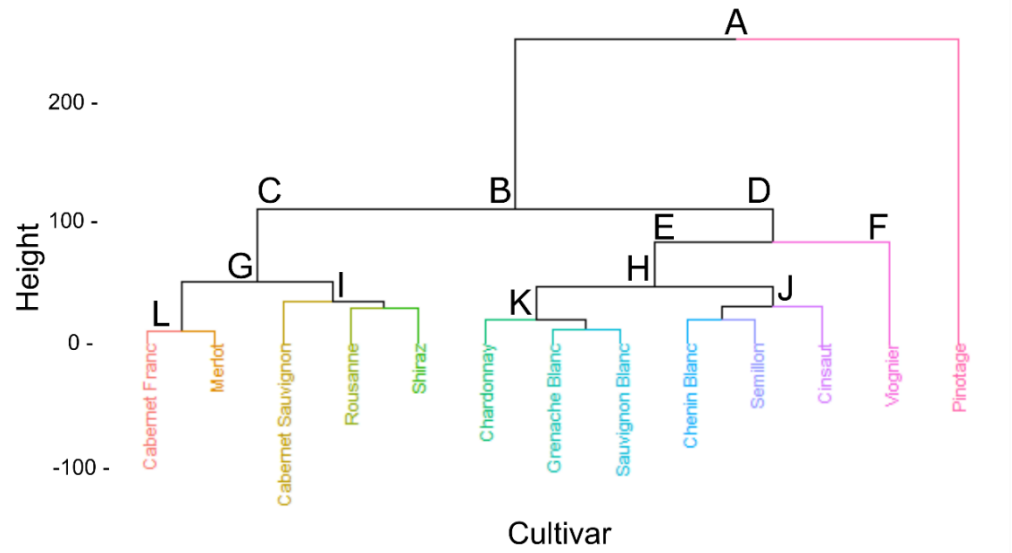


Figure 3.4. Dendrogram of the hierarchical agglomerative clustering analysis using the mean/median statistic and the average/centroid linkage method.

Thus, it can be hypothesised that although YAN concentration and composition are said to be affected by a multitude of extrinsic factors such as climate, soil, vintage, *etc.*, that cultivar may be an *overriding* factor in determining the concentration of YAN. This is presumed due to two reasons: Firstly, the association of cultivars based on their YAN concentrations and compositions mirrored that of their genetic relationships – thus, a strong genetic basis for YAN can be surmised. This shows the link between the genetics of the cultivar and its subsequent YAN concentration and composition.

Consequently, it is reasoned that YAN is cultivar-specific (being more similar when closer genetic relationships exist). Secondly, cultivar being nominated as an overriding factor is due to the fact that these relationships between cultivars (in terms of YAN concentration and composition) still mirrored the genetic relationships despite the fact that grape juice samples collected in this survey were collected from vineyards exposed to various extrinsic factors.

3.3.4. Relative importance of cultivar *vs.* district

A Classification and Regression Tree (CART) analysis was performed on the data. This technique is used to divide data into homogenous partitions with respect to the dependent variables. However, this technique is also useful in identifying the most important variables, in terms of explanatory power, in a particular dataset (Barlin *et al.*, 2013). In all cases, except for the FAN concentrations of white cultivars, cultivar was found to play a larger role than district in the 'decision' as to what the concentration of nitrogen (YAN, FAN or ammonia) will be (Figure 3.5). Moreover, studies investigating the effect of the environment on grapevine gene expression patterns (*i.e.* grapevine plasticity), found that genes involved in amino acid metabolism were less affected by the changing environmental conditions (Santo *et al.*, 2013). This is thought to be due to nitrogen metabolism forming part of the primary metabolism in grapevine and is in support of the hypothesis that cultivar plays an overriding role in determining the concentration and composition of YAN and that growing environment plays a subordinate role in the modulation of YAN.

The significant differences found between districts for certain white cultivars in the Games-Howell analysis seem to be because of the effect of district on FAN concentrations of these cultivars. This is because district is shown to play (although marginally) a bigger role in determining the concentration of FAN than cultivar for white cultivars. However, it must be kept in mind that, when having a closer look at the districts, it was found that district as demarcated by the Wine of Origin System may not have an effect *per se,* but rather the distance of the district from the ocean and subsequent air temperature may be a deciding factor of YAN concentration and composition.

As previously mentioned, CART analysis is able to partition data into homogenous groups. These partitions are based on the predictor variables with each partition being associated with a rule. Thus, the results of the CART analysis allowed the results of both of the 'tested' variables, cultivar and district, to be viewed simultaneously. As such, this analysis was able to visually highlight a few trends present in this particular dataset. Although statistically significant differences were not necessarily found between certain districts for certain cultivars, there were some districts which are found to repeatedly be associated with either higher or lower levels of YAN. The most striking examples are Franschhoek (n=46) and Stellenbosch (n=310) which are frequently observed to be associated with lower concentrations of this important component of grape juice, regardless of the cultivar. On the contrary, Paarl (n=171) and Swartland (n=56) are seen to frequently be associated with higher levels.

Breede River Valley region also seems to be associated with high YAN concentrations, however, only seven samples were collected from this region and, therefore, these results are less conclusive. Furthermore, certain cultivars were associated with either very high or very low concentrations of YAN, regardless of the district. These cultivars include Roussanne (n=15; very low) and Pinotage (n=12; very high). However, more samples are required to confirm these trends.

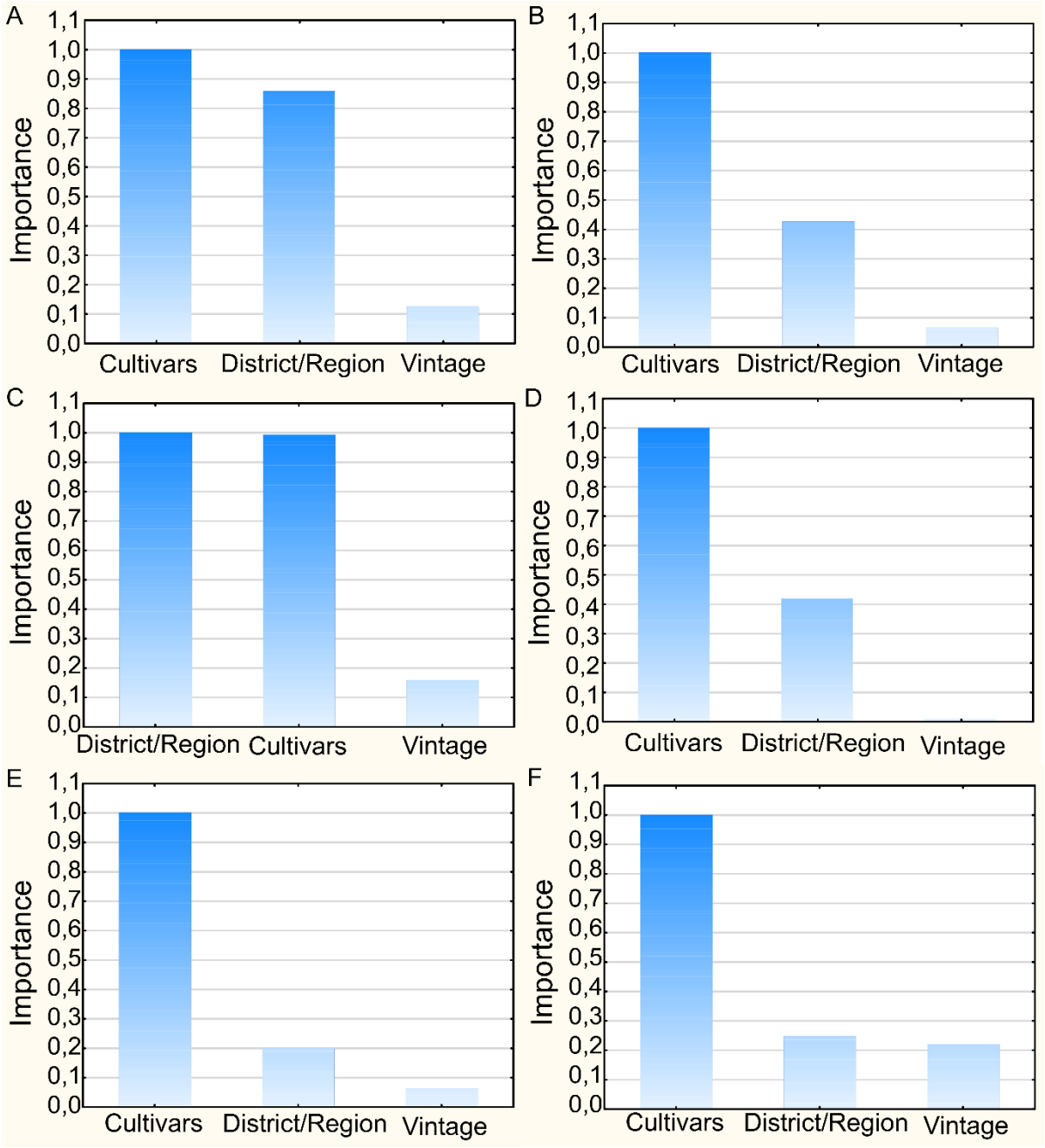


Figure 3.5. Importance plots generated through the CART analysis for YAN (A and B), FAN (C and D), and ammonia (E and F) concentrations for white (A, C and E) and red cultivars (B, D and F) investigated in this study.

As the importance plot has already indicated, cultivar has been identified as the most important deciding factor in the concentration and composition of YAN. This is illustrated in Figure 3.6 by certain cultivars always grouping together on either the lower or higher end of the plot regardless of the districts of origin. For example, Chenin Blanc and Sémillon feature only in box plot B and C, and Sauvignon Blanc, Grenache Blanc, Chardonnay, and Viognier feature only in box plot D and E. In other words, a cultivar such as Chenin Blanc, associated with lower levels of YAN is never found to group together with higher YAN-yielding cultivars such as Chardonnay and Sauvignon Blanc, regardless of the district. The same was true for higher YAN-yielding cultivars.

Although cultivar has been identified as the most important deciding factor for red cultivars as well, the box plots are not as clear-cut as they are for white cultivars (Figure 3.7). The reasons for this may be the closer genetic relationships exhibited by these red cultivars, and, subsequently, the more similar concentrations of YAN found between these cultivars in comparison to white cultivars included in this study. Furthermore, the smaller sample size of red cultivars compared to white may also impair the resolution of these results.

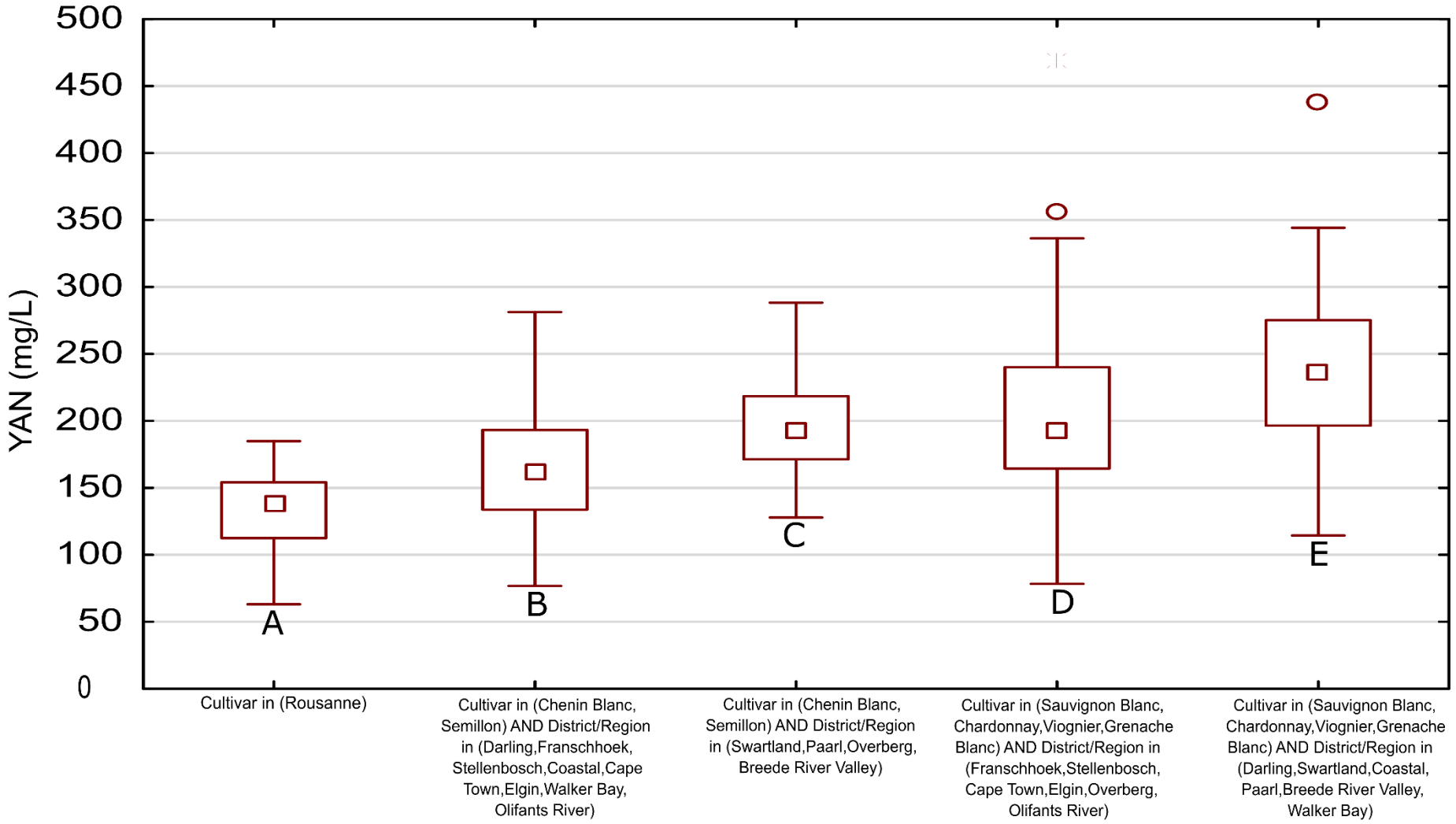


Figure 3.6. Box plots representing the results of the CART analysis for the YAN concentrations of white cultivars.

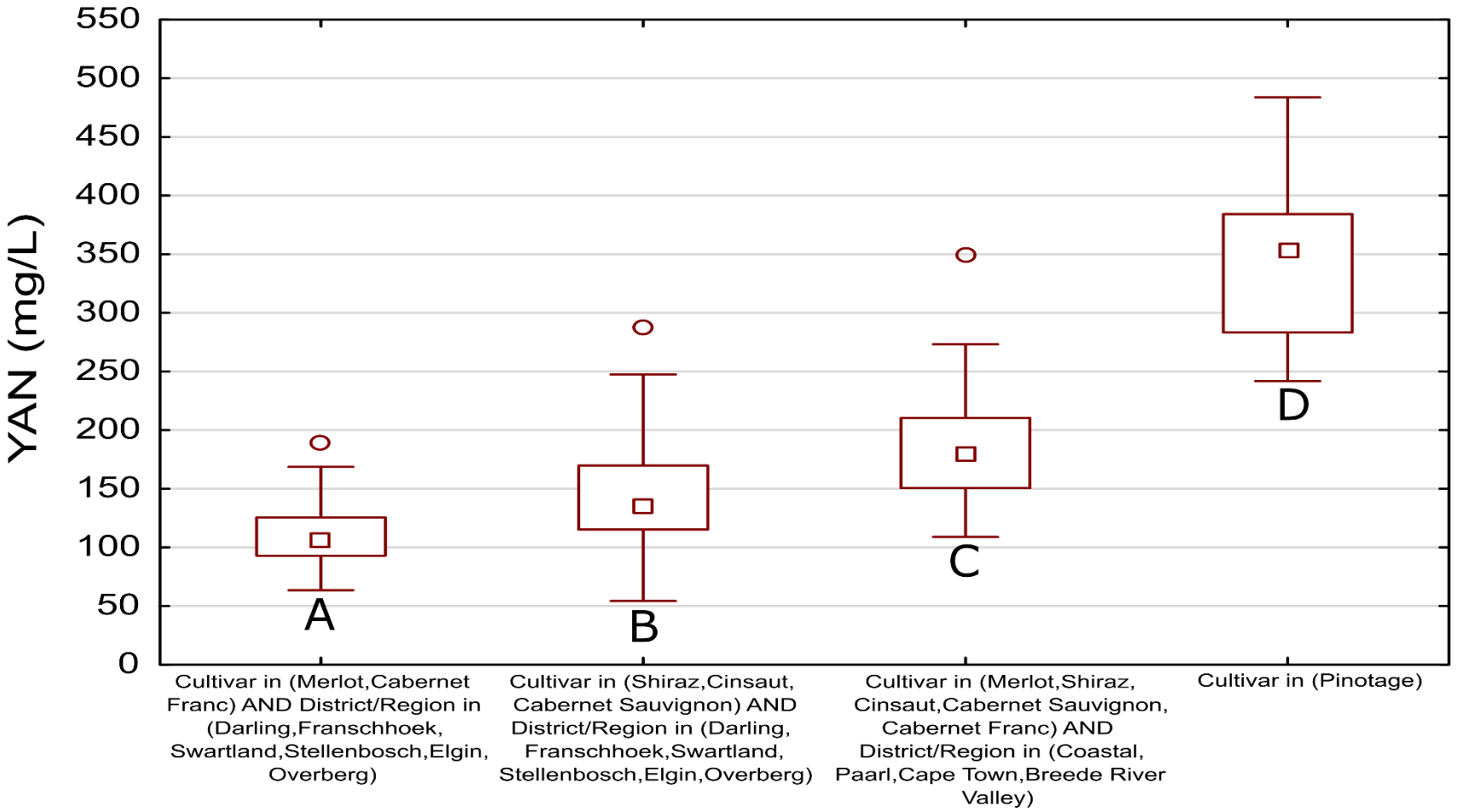


Figure 3.7. Box plots representing the results of the CART analysis for the YAN concentrations of red cultivars.

3.4. Conclusion

This study showed that YAN is indeed an inherent trait of a cultivar and although exact YAN concentrations cannot be predicted purely based on knowing the identity of the cultivar, a certain level of discernment is possible between some of the cultivars investigated. This was especially true between the red and white cultivars. Furthermore, this study highlighted which cultivars are most likely to require nitrogen additions to avoid the occurrence of stuck fermentations and those which could potentially run the risk of having excess nitrogen at the end of fermentation.

Aside from informing the local wine industry of the nitrogen status of some of South Africa's most important cultivars, this study aimed to lay the foundation for future experimental design for studies involving YAN. For example, in order to apply rapid and cost-effective methods such as infrared (IR) spectroscopy to predict YAN concentrations and to ensure the robustness of these prediction models, a representative sample set is required. However, what a 'representative' sample set is in the context of YAN requires an in-depth understanding of the nature of YAN. Based on the results of this survey, it was found that grape juice YAN concentration and composition is highly variable component of grape juice, covering an 11-fold range. Moreover, YAN concentrations showed increased variance with an increase in the number of samples. This indicates that the collection of a large number of samples is a particularly important factor in the context of YAN.

Furthermore, with the exception of Pinotage and Roussanne, the CA analysis showed a rather distinct separation in terms of the YAN concentrations of red and white cultivars, with red cultivars being associated with lower YAN concentrations compared to white cultivars. These findings may merit the separate consideration of red and white cultivars in YAN studies. However, as cultivars were found to be in most cases statistically significantly different from one another, hypothesised to be due to the difference of genes associated with nitrogen assimilation and metabolism between cultivars, there may be merit in not only considering red and white separately, but to consider *cultivars* separately. Subsequently, these results may indicate the need for cultivar-specific fertilization programmes to optimize nitrogen assimilation in the vineyard.

References

Ahad, N.A. & Yahaya, S.S.S., 2014. Sensitivity analysis of Welch's t-test. AIP Conference Proceedings 1605(1), 888–893.

Ahad, N.A., Othman, A.R., Yahaya, S.S.S., 2011. Comparative performance of pseudo-median procedure, Welch's test and Mann-Whitney-Wilcoxon at specific pairing. Modern Applied Science 5(5), 131–139.

Bailey, K.D., 1975. Cluster Analysis. Sociological Methodology 6(1), 59–128.

Bailey, K.D., 1983. Sociological classification and cluster analysis. Quality & Quantity 17(4), 251-268.

Barbosa, C., Mendes-Faia, A., Mendes-Ferreira, A., 2012. The nitrogen source impacts major volatile compounds released by Saccharomyces cerevisiae during alcoholic fermentation. International Journal of Food Microbiology 160(2), 87–93.

Barlin, J.N., Zhou, Q., et al., 2013. Classification and regression tree (CART) analysis of endometrial carcinoma: Seeing the forest for the trees. Gynecologic Oncology 130(3), 452–456.

Bell, S.-J. & Henschke, P.A., 2005. Implications of nitrogen nutrition for grapes, fermentation and wine. Australian Journal of Grape and Wine Research 11(3), 242–295.

Beltran, G., Esteve-Zarzoso, B., Rozès, N., Mas, A., Guillamón, J.M., 2005. Influence of the Timing of Nitrogen Additions during Synthetic Grape Must Fermentations on Fermentation Kinetics and Nitrogen Consumption. Journal of Agricultural and Food Chemistry 53(4), 996–1002.

Bely, M., Sablayrolles, J.-M., Barre, P., 1990. Automatic detection of assimilable nitrogen deficiencies during alcoholic fermentation in oenological conditions. Journal of Fermentation and Bioengineering 70(4), 246–252.

Berkhin, P., 2006. Nicholas, C., Kogan, J. (Eds) In: Grouping Multidimensional Data: Recent Advances in Clustering. Springer Berlin Heidelberg 25–71.

Betz, M.A., Gabriel, K.R., 1978. Type IV Errors and Analysis of Simple Effects. Journal of Educational and Behavioral Statistics 3(2), 121-143.

Bisson, L.F., 1999. Stuck and Sluggish Fermentations. Am J Enol Vitic. 50(1), 107-119.

Blashfield, R.K., 1976. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. Psychological Bulletin 83(3), 377–388.

Bonnardot, V., Planchon, O., Carey, V., Cautenet, S., 2002. Diurnal Wind, Relative Humidity and Temperature Variation in the Stellenbosch-Groot Drakenstein Wine-Growing Area. South African Journal of Enology and Viticulture 23(2), 62–71.

Box, G. & Anderson, L., 1955. Permutation Theory in the Derivation of Robust Criteria and the Study of Departures from Assumption. Journal of the Royal Statistical Society 17(1), 1–34.

Butzke, C.E., 1998. Survey of yeast assimilable nitrogen status in musts from California, Oregon, and Washington. American Journal of Enology and Viticulture 49(2), 220–224.

Carey, V.A., 2001. Spatial characterisation of natural terroir units for viticulture in the winegrowing area. Thesis, Stellenbosch University, Private Bag X1, 7602 Matieland (Stellenbosch), South Africa.

Christensen, P., 1984. Nutrient Level Comparisons of Leaf Petioles and Blades in Twenty-Six Grape Cultivars Over Three Years (1979 through 1981). American Journal of Enology and Viticulture 35(3), 124–133.

Coombs, W.T., Algina, J., et al., 1996. Univariate and Multivariate Omnibus Hypothesis Tests Selected to Control Type I Error Rates When Population Variances Are Not Necessarily Equal. Review of Educational Research 66(2), 137–179.

Dukes, B.C. & Butzke, C.E., 1998. Rapid Determination of Primary Amino Acids in Grape Juice Using an o-Phthaldialdehyde/N-Acetyl-L-Cysteine Spectrophotometric Assay. American Journal of Enology and Viticulture 49(2), 125–134.

Filipe-Ribeiro, L. & Mendes-Faia, A., 2007. Validation and comparison of analytical methods used to evaluate the nitrogen status of grape juice. Food Chemistry 100(3), 1272–1277.

Gishen, M., Dambergs, R.G., Cozzolino, D., 2005. Grape and wine analysis - enhancing the power of spectroscopy with chemometrics. Australian Journal of Grape and Wine Research 11(3), 296–305.

Gobbi, M., Comitini, F., D'Ignazi, G., Ciani, M., 2013. Effects of nutrient supplementation on fermentation kinetics, H2S evolution, and aroma profile in Verdicchio DOC wine production. European Food Research and Technology 236(1), 145–154.

Gump, B.H., Zoecklein, B.W., Fugelsang, K.C., Whiton, R.S., 2002. Comparison of Analytical Methods for Prediction of Pre-fermentation Nutritional Status of Grape Juice. Am. J. Enol. Vitic. 53(4), 325–329.

Hagen, K.M., Keller, M., Edwards, C.G., 2008. Survey of Biotin, Pantothenic acid, and Assimilable Nitrogen in Wine-grapes from the Pacific Northwest. Am. J. Enol. Vitic. 59(4), 432–436.

Harwell, M.R., Rubinstein, E.N., William S. Hayes, Corley, C., 1992. Summarizing Monte Carlo Results in Methodological Research: The One- and Two-Factor Fixed Effects ANOVA Cases. American Journal of Educational Statistics 17(4), 315–339.

Henschke, P.A. & Jiranek, V., 1993. Yeasts - metabolism of nitrogen compounds. In: Wine Microbiology and Biotechnology. Harwood Academic Publishers, Chur, Switzerland. pp. 77–164.

Hernandez-Orte, P., Bely, M., Cacho, J., Ferreira, V., 2006. Impact of ammonium additions on volatile acidity, ethanol, and aromatic compound production by different Saccharomyces cerevisiae strains during fermentation in controlled synthetic media. Australian Journal of Grape and Wine Research 12(2), 150–160.

Hildebrand, A.J., 2008. The Central Limit Theorem In: Math 408, Actuarial Statistics I.

Huang, Z. & Ough, C.S., 1989. Effect of Vineyard Locations, Varieties, and Rootstocks on the Juice Amino Acid Composition of Several Cultivars American. Journal of Enology and Viticulture 40(2), 135–139.

Levene, H., 1960. Robust tests for equality of variances. In: Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press.

Lieberman, M.D. & Cunningham, W.A., 2009. Type I and Type II error concerns in fMRI research: Re-balancing the scale. Social Cognitive and Affective Neuroscience 4(4), 423-428.

Lindman, H., 1992. Analysis of variance in experimental designs. Society for Industrial and Applied Mathematics 18(1), 134–137.

Mendes-Ferreira, A., Barbosa, C., Lage, P., Mendes-Faia, A., 2011. The Impact of Nitrogen on Yeast Fermentation and Wine Quality. Ciência Téc. Vitiv. 26(1), 17-32.

Moder, K., 2007. How to keep the Type I Error Rate in ANOVA if Variances are Heteroscedastic 2 Another View on the F-Ratio. Austrian Journal of Statistics 36(3), 179-188.

Murtagh, F. & Legendre, P., 2011. Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. Journal of Classification 31(3), 274-295.

Myburgh, P., 2005. Effect of altitude and distance from the Atlantic Ocean on mean February temperatures in the Western Cape Coastal region. Wynboer Technical Yearbook pp. 49–52.

Myles, S., Boyko, A.R., et al., 2011. Genetic structure and domestication history of the grape. Proceedings of the National Academy of Sciences 108(9), 3530–3535.

Nicolini, G., Larcher, R., et al., 2004. Status of yeast assimilable nitrogen in Italian grape musts and effects of variety, ripening and vintage. Vitis 43 (2), 89–96.

Nisbet, M.A., Martinson, T.E., Mansfield, A.K., 2014. Accumulation and Prediction of Yeast Assimilable Nitrogen in New York Wine-grape Cultivars. American Journal of Enology and Viticulture 65(3), 325–332.

Patz, C.D., Blieke, A., Ristow, R., Dietrich, H., 2004. Application of FT-MIR spectrometry in wine analysis. Analytica Chimica Acta 513(1), 81–89.

Pinotage association http://pinotage.co.za/. Accessed: April 2018.

Rusticus, S.A. & Lovato, C.Y., 2014. Impact of Sample Size and Variability on the Power and Type I Error Rates of Equivalence Tests: A Simulation Study. Practical Assessment, Research & Evaluation 19(11), 1–10.

Saayman, D., 1999. The development of vineyard zonation and demarcation. In: South Africa WineLand pp. 2–5.

Santo, S.D., Tornielli, G.B., Zenoni, S., Fasoli, M., Farina, L., Anesi, A., Guzzo, F., Delledonne, M., Pezzotti, M., 2013. The plasticity of the grapevine berry transcriptome. Genome Biology 14(6), 1-18.

Saraçli, S., Doğan, N., Doğan, I., 2013. Comparison of hierarchical cluster analysis methods by cophenetic correlation. Journal of Inequalities and Applications 2013(203), 1–8.

SAWIS, 2018. Production areas: Districts. South African Wine Industry Information and Systems. Paarl. http://www.sawis.co.za/cert/productionareas.php?search. Accessed: April 2018.

Shah, N., Cynkar, W., Smith, P., Cozzolino, D., 2010. Use of Attenuated Total Reflectance Midinfrared for Rapid and Real-Time Analysis of Compositional Parameters in Commercial White Grape Juice. Journal of Agricultural and Food Chemistry 58(6), 3279–3283.

Shapiro, S.S., Wilk, M.B., 1965. An Analysis of Variance Test for Normality. Biometrika 52(3/4), 591–611.

Sokal, R.R. & Rohlf, F.J., 1962. The Comparison of Dendrograms by Objective Methods. Taxon 11(2), 33-40.

Sokal, R.R. & Sneath, P.H.A., 1963.Principals of Numerical Taxonomy. W.H. Freeman & Co., New York.

Southey, T.O., 2017. Integrating climate and satellite remote sensing to assess the reaction of Vitis vinifera L. Cv. Cabernet Sauvignon to a changing environment. Thesis, Stellenbosch University, Private Bag X1, 7602 Matieland (Stellenbosch), South Africa.

Vink, N., Deloire, A., Bonnardot, V. & Ewert, J., 2010. Terroir, climate change, and the future of South Africa's wine industry. International Journal of Climate Change Strategies and Management 4(4), 420-441.

Appendix A

Chapter 3
Additional Tables and Figures

Chapter 3 Appendix A: Additional Tables and Figures

Table A3.1. Descriptive statistics of YAN, FAN, and ammonia per cultivar (mg N/L).

	Number of samples	Average	Min	Lower Quartile	Median	Upper Quartile	Max
YAN							
Chardonnay	97	223±57	115	181	218	261	469
Chenin Blanc	175	176±43	77	147	176	205	288
Grenache Blanc	17	213±58	122	171	198	242	330
Roussanne	15	132±34	63	119	139	152	185
Sauvignon Blanc	221	208±56	78	169	201	247	438
Semillon	16	180±51	95	140	185	218	265
Viognier	42	250±56	151	210	253	275	438
Cabernet Franc	13	108±26	64	95	107	113	169
Cabernet Sauvignon	39	146±45	55	116	138	170	273
Cinsaut	15	194±63	112	140	187	237	348

Table A3.1. (cont.)

	Number of samples	Average	Min	Lower Quartile	Median	Upper Quartile	Max
Merlot	30	110±26	65	93	109	126	190
Pinotage	12	348±77	242	284	354	384	484
Shiraz	50	151±43	54	120	148	178	288
FAN							
Chardonnay	97	155±41	78	126	154	178	302
Chenin Blanc	175	133±32	64	112	132	157	221
Grenache Blanc	17	143±48	82	102	146	171	251
Roussanne	15	106±27	55	92	114	124	144
Sauvignon Blanc	221	146±42	59	116	138	172	341
Semillon	16	123±38	67	89	123	145	188
Viognier	42	188±44	104	158	185	210	315
Cabernet Franc	13	90±19	55	80	90	99	131
Cabernet Sauvignon	39	99±23	48	87	95	110	158
Cinsaut	15	125±54	30	89	119	158	243

Table A3.1. (cont.)

	Number of samples	Average	Min	Lower Quartile	Median	Upper Quartile	Max
Merlot	30	86±18	55	75	82	94	140
Pinotage	12	249±60	157	217	238	265	365
Shiraz	50	123±36	42	97	123	144	239
Ammonia							
Chardonnay	97	67±22	16	54	64	79	167
Chenin Blanc	175	43±15	9	31	42	53	86
Grenache Blanc	17	70±17	40	60	69	79	100
Roussanne	15	25±11	9	16	28	32	41
Sauvignon Blanc	221	62±19	14	48	62	73	115
Semillon	16	57±16	27	52	57	68	77
Viognier	42	63±15	39	51	61	72	122
Cabernet Franc	13	17±10	6	10	14	22	38
Cabernet Sauvignon	39	47±26	1	28	41	68	127
Cinsaut	15	69±16	47	57	73	79	105

Table A3.1. (cont.)

	Number of samples	Average	Min	Lower Quartile	Median	Upper Quartile	Max
Merlot	30	24±10	7	18	23	29	50
Pinotage	12	99±25	62	82	105	113	140
Shiraz	50	28±11	11	19	25	32	56

Table A3.2. Random sample sizes per cultivar of 10%, 50% as well as 100% of the data and the corresponding variance and standard deviations.

Cabernet Sauvignon		**YAN**	**FAN**	**Ammonia**
10%	Variance	290	70	227
	Standard Deviation	17	8	15
50%	Variance	2032	437	451
	Standard Deviation	45	21	21
100%	Variance	2042	527	662
	Standard Deviation	45	23	26
Chardonnay		**YAN**	**FAN**	**Ammonia**
10%	Variance	2216	954	451
	Standard Deviation	47	31	21
50%	Variance	2770	1374	350
	Standard Deviation	53	37	19
100%	Variance	2639	1459	392
	Standard Deviation	51	38	20
Chenin Blanc		**YAN**	**FAN**	**Ammonia**
10%	Variance	1393	734	221
	Standard Deviation	37	27	15
50%	Variance	1687	1052	225
	Standard Deviation	41	32	15
100%	Variance	1843	1036	230
	Standard Deviation	43	32	15
Merlot		**YAN**	**FAN**	**Ammonia**
10%	Variance	742	386	82
	Standard Deviation	27	20	9
50%	Variance	620	347	80
	Standard Deviation	25	19	9
100%	Variance	688	337	104
	Standard Deviation	26	18	10
Sauvignon Blanc		**YAN**	**FAN**	**Ammonia**
10%	Variance	2709	1486	333
	Standard Deviation	52	39	18
50%	Variance	2977	1641	359
	Standard Deviation	55	41	19
100%	Variance	3088	1735	351
	Standard Deviation	56	42	19
Shiraz		**YAN**	**FAN**	**Ammonia**
10%	Variance	55	23	15
	Standard Deviation	7	5	4
50%	Variance	1697	1167	100
	Standard Deviation	41	34	10
100%	Variance	1851	1272	128
	Standard Deviation	43	36	11

Table A3.2 (cont.)

Viognier		*YAN*	*FAN*	*Ammonia*
10%	Variance	1707	1110	139
	Standard Deviation	41	33	12
50%	Variance	2249	1667	121
	Standard Deviation	47	41	11
100%	Variance	3042	1917	241
	Standard Deviation	55	44	16

Table A3.3. Games-Howell post-hoc test to test significant differences between cultivars contributing >30 samples. Values indicated in red are significant at $p<0.05$.

		Cabernet Sauvignon	Chardonnay	Chenin Blanc	Merlot	Sauvignon Blanc	Shiraz	Viognier
YAN	**Cabernet Sauvignon**							
	Chardonnay	0,000						
	Chenin Blanc	0,007	0,000					
	Merlot	0,002	0,000	0,000				
	Sauvignon Blanc	0,000	0,478	0,000	0,000			
	Shiraz	0,999	0,000	0,006	0,000	0,000		
	Viognier	0,000	0,057	0,000	0,000	0,001	0,000	
FAN	**Cabernet Sauvignon**							
	Chardonnay	0,000						
	Chenin Blanc	0,000	0,000					
	Merlot	0,115	0,000	0,000				
	Sauvignon Blanc	0,000	0,680	0,011	0,000			
	Shiraz	0,004	0,000	0,524	0,000	0,002		
	Viognier	0,000	0,001	0,000	0,000	0,000	0,000	

Table A3.3. (cont.)

		Cabernet Sauvignon	Chardonnay	Chenin Blanc	Merlot	Sauvignon Blanc	Shiraz	Viognier
Ammonia	**Cabernet Sauvignon**							
	Chardonnay	0,002						
	Chenin Blanc	0,952	0,000					
	Merlot	0,000	0,000	0,000				
	Sauvignon Blanc	0,026	0,466	0,000	0,000			
	Shiraz	0,001	0,000	0,000	0,860	0,000		
	Viognier	0,045	0,822	0,000	0,000	1,000	0,000	

Table A3.4. Games-Howell post-hoc test to test significant differences between districts contributing more than 5 samples of Cabernet Sauvignon grape juices.

Cabernet Sauvignon		Darling	Franschhoek	Stellenbosch
YAN	Darling			
	Franschhoek	0.901		
	Stellenbosch	0.883	0.981	
FAN	Darling			
	Franschhoek	0.470		
	Stellenbosch	0.694	0.708	
Ammonia	Darling			
	Franschhoek	0.975		
	Stellenbosch	0.986	0.943	

Table A3.5. Games-Howell post-hoc test to test significant differences between districts contributing more than 5 samples of Merlot grape juices.

Merlot		Darling	Franschhoek	Stellenbosch
YAN	Darling			
	Franschhoek	0.198		
	Stellenbosch	0.383	0.697	
FAN	Darling			
	Franschhoek	0.170		
	Stellenbosch	0.361	0.552	
Ammonia	Darling			
	Franschhoek	0.424		
	Stellenbosch	0.563	0.916	

Table A3.6. Games-Howell post-hoc test to test significant differences between districts contributing more than 5 samples of Shiraz grape juices

Shiraz		Darling	Paarl	Stellenbosch	Swartland
YAN	Darling				
	Paarl	0.082			
	Stellenbosch	0.295	0.756		
	Swartland	0.930	0.440	0.826	
FAN	Darling				
	Paarl	0.025			
	Stellenbosch	0.127	0.857		
	Swartland	0.843	0.557	0.867	
Ammonia	Darling				
	Paarl	0.623			
	Stellenbosch	0.903	0.791		
	Swartland	1.000	0.345	0.676	

Table A3.7. Games-Howell post-hoc test to test significant differences between districts contributing more than 5 samples of Chardonnay grape juices.

Chardonnay		Elgin	Paarl	Stellenbosch
YAN	Elgin			
	Paarl	0.000		
	Stellenbosch	0.001	0.211	
FAN	Elgin			
	Paarl	0.001		
	Stellenbosch	0.006	0.308	
Ammonia	Elgin			
	Paarl	0.006		
	Stellenbosch	0.020	0.296	

Table A3.8. Games-Howell post-hoc test to test significant differences between districts contributing more than 5 samples of Chenin Blanc grape juices.

Chenin Blanc		Coastal	Olifants River	Paarl	Stellenbosch	Swartland	Walker Bay
YAN	Coastal						
	Olifants River	0.999					
	Paarl	0.254	0.719				
	Stellenbosch	0.991	1.000	0.045			
	Swartland	0.287	0.747	1.000	0.108		
	Walker Bay	0.631	0.565	0.075	0.346	0.078	
FAN	Coastal						
	Olifants River	0.943					
	Paarl	0.581	0.293				
	Stellenbosch	1.000	0.817	0.032			
	Swartland	0.634	0.313	1.000	0.114		
	Walker Bay	0.589	0.980	0.072	0.346	0.077	
Ammonia	Coastal						
	Olifants River	0.012					
	Paarl	0.035	0.228				
	Stellenbosch	0.563	0.046	0.455			
	Swartland	0.044	0.262	1.000	0.480		
	Walker Bay	0.939	0.015	0.132	0.463	0.128	

Table A3.9. Games-Howell post-hoc test to test significant differences between districts contributing more than 5 samples of Sauvignon Blanc grape juices

Cultivar:	Sauvignon Blanc	Cape Town	Darling	Elgin	Franschhoek	Olifants River	Overberg	Paarl	Stellenbosch	Swartland	Walker Bay
YAN	Cape Town										
	Darling	0.740									
	Elgin	0.820	0.001								
	Franschhoek	1.000	0.871	0.429							
	Olifants River	0.997	0.817	1.000	0.986						
	Overberg	0.998	0.426	1.000	0.974	1.000					
	Paarl	0.840	1.000	0.009	0.938	0.846	0.512				
	Stellenbosch	1.000	0.295	0.180	1.000	0.994	0.991	0.512			
	Swartland	0.048	0.217	0.025	0.072	0.209	0.026	0.202	0.056		
	Walker Bay	0.008	0.102	0.000	0.015	0.262	0.010	0.099	0.002	1.000	
FAN	Cape Town										
	Darling	0.773									
	Elgin	0.336	0.000								
	Franschhoek	1.000	0.403	0.388							
	Olifants River	0.989	0.713	1.000	0.996						
	Overberg	0.994	0.520	1.000	0.999	1.000					
	Paarl	0.998	0.986	0.015	0.942	0.907	0.869				
	Stellenbosch	1.000	0.122	0.092	1.000	0.997	0.999	0.777			
	Swartland	0.132	0.387	0.048	0.106	0.155	0.072	0.215	0.102		
	Walker Bay	0.050	0.422	0.001	0.023	0.216	0.049	0.122	0.013	0.999	
Ammonia	Cape Town										
	Darling	0.835									
	Elgin	1.000	0.816								
	Franschhoek	0.812	1.000	0.843							
	Olifants River	1.000	0.964	0.999	0.935						

Table A3.9. (cont.)

Cultivar:	Sauvignon Blanc	Cape Town	Darling	Elgin	Franschhoek	Olifants River	Overberg	Paarl	Stellenbosch	Swartland	Walker Bay
	Overberg	1.000	0.513	1.000	0.627	1.000					
	Paarl	0.195	0.689	0.191	0.989	0.731	0.083				
	Stellenbosch	0.965	1.000	0.977	0.996	0.984	0.807	0.510			
	Swartland	0.404	0.736	0.444	0.933	0.645	0.338	1.000	0.649		
	Walker Bay	0.009	0.048	0.012	0.446	0.434	0.004	0.943	0.026	1.000	

Table A3.10. Cophenetic correlation coefficient for the various statistics and clustering methods tested. The conditional formatting indicates the lowest to highest correlations.

Clustering Method	Interquartile Range	Average	Median
Single	0.5485235	0.8362645	0.8696419
Complete	0.7200294	0.8429283	0.8401101
Average	0.7274925	0.8770316	0.8856007
Ward	0.7120145	0.5489079	0.5497238
Median	0.7157945	0.867917	0.8658887
Centroid	0.7121264	0.8656678	0.8888572

Low High

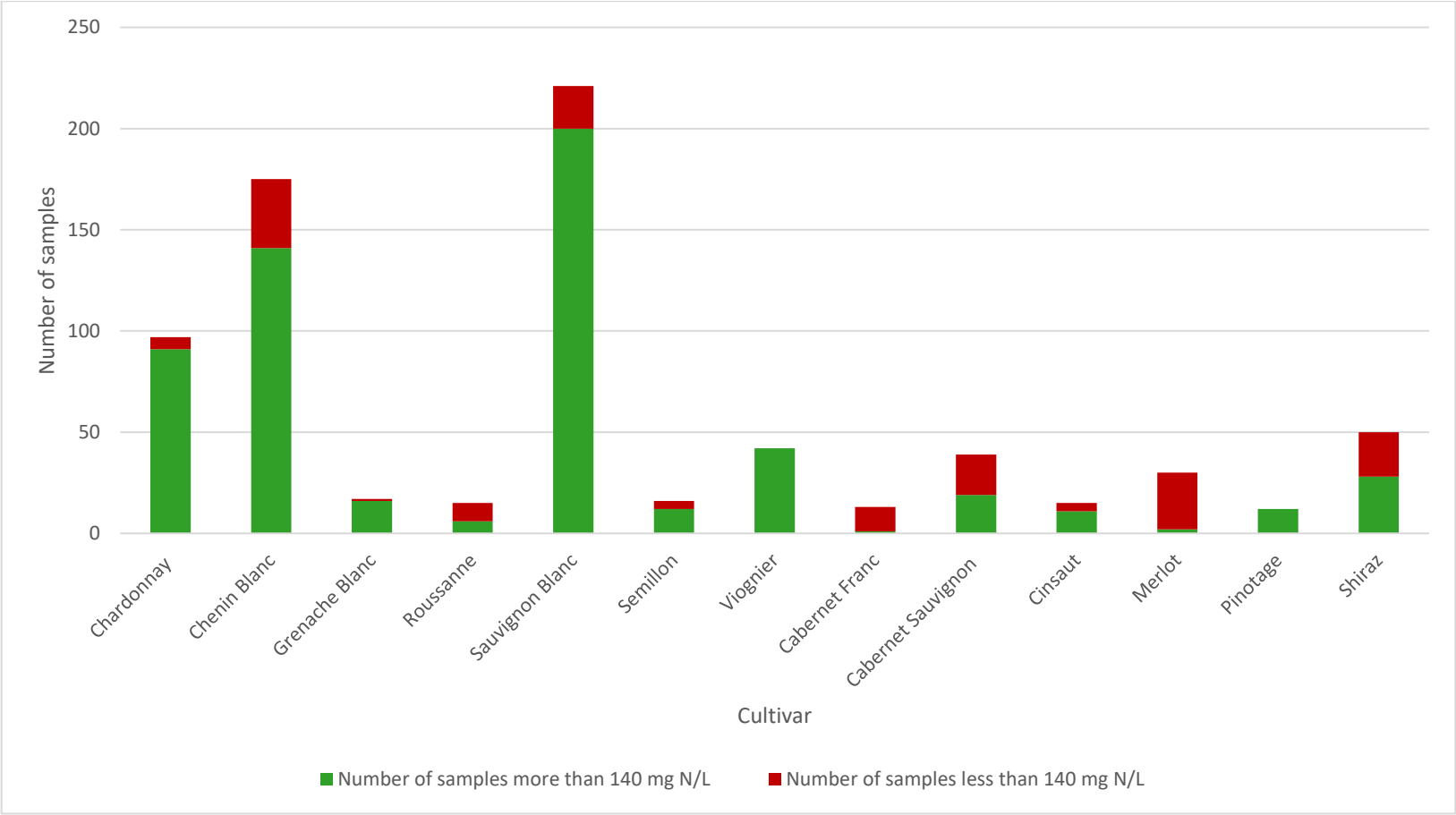


Figure A3.1. Number of samples collected during the 2016 and 2017 that were found to be above and below the recommended level of 140 mg N/L of YAN.

Chapter 4

Research Results

Viability of IR Spectroscopy for the Accurate Measurement of N Content of Grape Juice

Chapter 4
Viability of IR Spectroscopy for the Accurate Measurement of N Content of Grape Juice

4.1. Introduction

Infrared (IR) energy forms part of the electromagnetic spectrum between the visible and microwave region and can be divided into the near infrared (NIR) and the mid-infrared region (MIR), which corresponds to the wavelengths 780-2500 nm (wavenumbers 13400-4000 cm^{-1}) and 2500-25000 nm (wavenumbers 4000-400 cm^{-1}), respectively (Cozzolino, 2009). IR spectroscopy, also known as vibrational spectroscopy, is able to quantify the biochemical compounds present in grape (juice) and wine due to the fundamental molecular vibrations caused by the various functional groups present in a sample. As such, when a sample is exposed to IR light, it will absorb the wavelengths of light which match the vibrations of the specific functional groups present, while the light of other frequencies will either be transmitted or reflected. This can subsequently be used to infer the concentration of the specific compound present in the sample (Bauer *et al.*, 2008; Gishen *et al.*, 2010; Cozzolino *et al.*, 2011).

Thus, this technology provides the possibility of "fingerprinting" samples and, therefore, can provide an in-depth understanding of the chemical properties of various food and beverage products (Gishen *et al.,* 2010). However, the potential of spectroscopic techniques would not have been realised if it had not been for the major developments in the field of chemometrics. Chemometric techniques such as partial least squares (PLS) regression and principal component regression (PCR) allow the simultaneous consideration of multiple variables and are also able to handle highly correlated and noisy data, addressing the inherent issues related to dealing with spectroscopic data (Cozzolino *et al.*, 2009). This is due to the fact that these techniques extract latent variables from the original spectral data, thereby reducing the number of X-variables (spectral data points) to a set of non-correlated variables. This set of non-correlated variables can then be used to explain the variation in the data (Wold *et al.*, 2001; Liu *et al.*, 2011) and subsequently, provide the possibility of building suitable and robust calibration models.

Due to the complexity of the winemaking process and the increasing consumer demand for high quality wines, monitoring grape and wine composition has become a necessity (Gishen *et al.*, 2010). However, timely and cost-effective analysis is not always possible using conventional methods. This is owed to the fact that often these methods cannot be carried out on-site as they require trained personnel and the use of potentially hazardous chemicals (Gump *et al.*, 2002). Thus, the possibility of providing simple, rapid and cost-effective methods which are non-destructive and environmentally friendly would be an indispensable asset to the modern wine industry. These properties are all

characteristic of IR spectroscopy, and although there has been widespread adoption of this technology in the food industry, the use of IR spectroscopy in the wine industry is still in its infancy (Bauer *et al.*, 2008; Martelo-Vidal & Vázquez, 2014; Cozzolino, 2015).

The possible reasons for this has been highlighted by Dambergs, Gishen, and Cozzolino (2015). The most pertinent being the lack of understanding of the technology. This is because it hampers the possibility of building robust calibrations, capable of providing accurate results for samples which are: (i) exposed to different environmental conditions, (ii) from different varieties and (iii) from different vintages (Nicolaï *et al.*, 2007). These are essential factors to consider for the successful integration of this technology into the wine industry, especially due to the notoriously complex nature of the grape juice matrix (Bauer *et al.*, 2008). As a result, obtaining a representative calibration set becomes a particularly challenging task (Patz *et al.*, 2004). Furthermore, the bulk of publications currently available on spectroscopic modelling in grape and wine research generally use a limited sample set and thus, chances are that the large degree of the variation naturally present in the population is neglected (Skoutelas *et al.*, 2011; Dambergs *et al.*, 2015). Moreover, more often than not, these publications also do not test their models using independent validation sets but rather report values for cross-validation which are in most cases, overoptimistic (Versari *et al.*, 2008; Shah *et al.*, 2010). Cross-validation (CV) entails splitting the sample set into a predetermined number of subsets. Calibrations are then obtained by removing a different subset from the calibration data until each subset has been left out once. Thus, CV may lead to overoptimistic results as the samples used to validate the model have also been used to calibrate the model (Anderssen *et al.*, 2006).

To date, the bulk of the research has focused on building calibration models for compositional parameters of intact grapes, skins and seeds as well as wine (Dambergs *et al.*, 2015). Furthermore, these studies focused on phenolic compounds and other indicators of ripening such as total soluble solids (TSS) total acidity (TA) and Brix as well as the ethanol content of finished wines (Patz *et al.*, 2004; Larrain *et al.*, 2008; Schaare *et al.*, 2012; Cozzolino, 2015). However, there is a lack of studies reporting calibrations for compounds present in grape juice or must right before or during the winemaking process.

Yeast assimilable nitrogen (YAN) can be defined as nitrogen sources present in the grape juice matrix that can be taken up by yeast during fermentation. These sources include free amino nitrogen (FAN) and ammonia (Bell & Henschke, 2005). YAN is an essential component of grape juice as it plays a major role in fermentation efficiency by providing the necessary nutrients required for the growth and proliferation of yeast, thereby reducing the chances of stuck or sluggish fermentations (Henschke & Jiranek, 1993). Furthermore, YAN has been highlighted as a driver of quality by influencing the organoleptic qualities of wine (Ugliano *et al.*, 2007). This is primarily owed to the free amino nitrogen (FAN) portion of YAN, as certain amino acids (branched-chain and aromatic amino acids) have been identified as precursor molecules for the production of particular aroma compounds

(Smit, 2013). Thus, it is important to not only measure total YAN before the start of fermentation, but also to have knowledge of the composition. Consequently, this information will ensure more informed decision-making regarding nutrient supplementation strategies and assist in avoiding unnecessary prophylactic nutrient additions.

Up until now, there have been only a few reports on the measurement of YAN and/or its components using IR spectroscopy. The earliest report was done by Manley, van Zyl and Wolf (2001). This study attempted to calibrate an FT-NIR instrument for the measurement of FAN, using 97 settled grape juice samples from various white varieties. They were, however, unsuccessful, obtaining a large standard error of prediction (SEP) of 272.1 mg N/L. Thus, instead, a Soft Independent Modelling by Class Analogy (SIMCA) was used to classify the samples as having either high, medium or low concentrations of FAN. In a comparison done by Dambergs *et al.* (2004), MIR was shown to outperform NIR for the measurement of YAN, FAN, and ammonia, as higher ratio of standard error of performance to standard deviation (RPD) and lower standard error of cross-validation (SECV) values were observed using MIR. On the other hand, Shah *et al*. (2010), investigated the viability of using ATR-MIR to measure various grape juice parameters including YAN, FAN, and ammonia. SEP values of 42.4 mg N/L, 36.7 mg N/L and 17.2 mg N/L were obtained for YAN, FAN, and ammonia, respectively. Furthermore, a RPD of approximately 2 was obtained for each of these parameters, indicating a qualitative rather than quantitative determination of these grape juice parameters. In another study, 71 grape juice samples from the Lisbon region in Portugal were used to build a calibration for YAN using FT-MIR spectroscopy. An R^2 of 0.993, SEP of 5.9 mg N/L and an RPD of 7.8 was obtained (Skoutelas *et al.*, 2011). These results may, however, be overoptimistic due to the limited number of samples included in the model in combination with the use of a cross-validation strategy rather than external validation.

Thus, IR spectroscopy shows potential for the measurement of YAN concentration and composition. However, for this technology to become a feasible option for industry, a few key issues need to be addressed. These include building calibrations with larger data sets including different varieties, origins, and vintages, as well as independent validation to adequately test the accuracy and robustness of these models. Therefore, the aim of this study is to fully investigate the viability of various infrared spectroscopic instruments for the accurate quantification of YAN, FAN, and ammonia concentrations by incorporating independent and robust validation strategies.

4.2. Materials and Methods

4.2.1. Sample collection

A total of 905 grape juice samples were collected over three vintages (2016 – 2018). Samples were collected directly from commercial wineries at a ripeness level suitable for commercial winemaking. Red grape juice samples were collected after crushing and white after settling. To increase the chances of obtaining a representative sample set of the South African wine industry, an unsupervised strategy was employed. This meant that no specific cultivars or origin was targeted. Consequently, samples were collected from 28 different cultivars, of which 12 were white and 16 were red. Furthermore, these samples were collected from 14 different grape-growing districts situated in the Western Cape of South Africa. These grape-growing districts were classified according to the demarcation set by the Wine of Origin System of South Africa (SAWIS, 2017). Samples were coded immediately upon collection and stored at -20°C until analysis.

4.2.2. Analytical Methods

4.2.2.1. Reference Method

The components of YAN, FAN and ammonia, were measured separately by enzymatic assay using the Megazyme™ K-PANOPA (Ireland) for FAN and Enzytec™ Fluid Ammonia (R-Biopharm, Germany) for ammonia. This was performed on the Arena 20XT (Thermo Fisher Scientific, Waltham, MA) which provides automated spectrophotometric readings. The individual values for FAN and ammonia were then summed to determine the total amount of YAN available and were expressed as mg N/L.

4.2.2.2. Infrared spectroscopy scanning

The samples were thawed at room temperature on the day of analysis and were centrifuged at 5000 rpm for 5 min in a 7366 Hermle centrifuge (Wehingen, Germany) prior to analysis. Spectra were collected from three bench-top infrared instruments, namely: a multi-purpose analyser (MPA) FT-NIR instrument (Bruker Optics, Ettlingen, Germany), Alpha-P ATR FT-MIR spectrometer (Bruker Optics, Ettlingen, Germany), and WineScan™ FT120 (FOSS Electric, Denmark).

FT-NIR spectra (12500-4000 cm^{-1}) were collected by the MPA in transmission by placing samples in a 1 mm cuvette. The absorbance spectrum obtained for each sample was acquired at a resolution of 2 cm^{-1} and at a scanning velocity of 10 kHz, averaged over 32 scans. Air was used as background and an air spectrum was taken periodically during the scanning of the samples and was automatically subtracted from each individual sample spectrum.

Spectra in the mid-infrared range (4000-600 cm^{-1}) were collected by the Alpha-P ATR FT-MIR spectrometer. Each sample was scanned at a resolution of 4 cm^{-1} and at a scanning velocity of 7.5 kHz, averaged over 64 scans to give a final reading. Instrumental control of the MPA FT-NIR and the Alpha-P ATR FT-MIR were carried out using OPUS software (OPUS v. 7.0 for Microsoft, Bruker Optics, Ettlingen, Germany).

The WineScan™ FT120 measures primarily in the mid-infrared region (4000-929 cm^{-1}), however, a small section of the near-infrared region is also included (5011-4000 cm^{-1}). This instrument recorded spectra at a resolution of 4 cm^{-1} in transmission which was then converted into a linearized absorbance spectrum. Each measurement was averaged over 20 readings to give a final measurement. Prior to analysis of the grape juice samples, the background absorbance in the grape juice sample is accounted for using the FOSS Zero Liquid S-6060 (WineScan™ manual).

4.2.3. Data Analysis

Calibration models and model accuracy were evaluated using OPUS software (OPUS v. 7.2 for Microsoft, Bruker Optics, Ettlingen, Germany). This software correlates the reference values to the spectra through the use of the partial least-squares (PLS) regression algorithm. The accuracy and reliability of the models were assessed based on a set of performance evaluation indices which included the coefficient of determination for calibration and validation ($R^2_{CAL}$and R^2_{VAL}), the root-mean square error of calibration (RMSEC) and validation (RMSEP) as well as the RPD in calibration and validation (RPD_{CAL} and RPD_{VAL}).

The coefficient of determination, R^2_{CAL} and R^2_{VAL}, is a measure of how well the variation observed in the predictor variables can be explained by the response variables in the calibration and validation set, respectively. This value ranges between 0 and 1, where 0 explains none of the variation and 1 the total amount of variation present. Thus, the closer the R^2-value is to 1, the more variation can be explained and accounted for by the model (Bauer *et al.*, 2008; Aleixandre-Tudo *et al.*, 2018).

Furthermore, the RMSEC and RMSEP measures the mean difference between the concentration values obtained from the reference method and the values predicted by the model in the calibration and validation steps, respectively. In other words, this value represents the average uncertainty that is anticipated for the prediction of new samples (Nicolaï *et al.*, 2007; Aleixandre-Tudo *et al.*, 2018).

The RPD of a model indicates the ratio of the standard deviation of either the calibration or validation set to the corresponding error in prediction (RMSEC or RMSEP) (RPD = SD/RMSE). Thus, the more variability accounted for by the model (SD) and the lower the error in prediction (RMSEP), the higher the RPD will be, and subsequently, the more reliable the model is. RPD values ranging between 1.5 and 2 are considered only sufficient to discriminate high values from low, whereas values between 2 and 2.5 allows for coarse quantification. A value above 2.5 indicates a good level of quantification,

however, values above 3 are preferred (Nicolaï *et al.*, 2007). When RPD values reach 5 or above, these models are thought to be viable for quality control (Shah *et al.*, 2010).

The optimum number of latent variables (*i.e.* rank) to avoid overfitting of the model was algorithmically determined as described by Aleixandre-Tudo *et al.* (2018), Rank was, however, not used as a criteria to compare the reliability of the models in this study. Instead, a provision was made which allowed for a maximum of 20 latent variables to be considered during model optimization. This number was considered to be low enough to avoid overfitting of the models as YAN is a minor component, producing a rather weak signal in a highly complex matrix. Moreover, the chances of overfitting were further decreased by external validation strategies in addition to the large number of samples that were gathered from a variety of different cultivars, vintages and origins – ensuring that both calibration and validation sets would be representative of the population.

The following strategy was employed during the modelling process: For each instrument, the spectra from all three vintages (which included all the different varieties and origins of grape juice samples) were uploaded to the OPUS software with their corresponding reference values for either YAN, FAN, or ammonia. The sample set was divided into a 66/34 calibration to validation set using the Kennard-Stone algorithm for random selection by selecting the “automatic selection of test samples” feature. Thus, an external validation set was used to validate the models. The models were then let to run using the “general B” option incorporated in the software package. This option automatically divides the spectra into 10 sub-regions. The regions used for the top 5 models were further investigated for optimization of the calibration model. These regions were then manually selected using the “user defined optimization regions” function which allows a manual selection of 10 sub-regions of any size using the “general B” option. Furthermore, pre-processing techniques such as smoothing, standardization, transformation, and normalization were used for model optimization.

Once the optimum regions were identified for a specific instrument and sample parameter, a subsequent model was built using these settings, but instead the sample set was divided into a 50/50 ratio of calibration/test. The models including samples from all the different varieties, origins and vintages will from hereon be referred to as global models and differentiated based on their calibration to validation ratio (66/34 or 50/50).

During the optimization of each model, outliers were removed and the pre-processing method which resulted in the lowest RMSEP and highest RPD was selected. Outliers were detected by the Mahalanobis distance. The Mahalanobis distance is calculated for each calibration spectrum from which a threshold is calculated. This threshold determines whether the spectra of an unknown sample can be reliably predicted or not.

To assess the robustness of the models, it was tested to see whether the YAN, FAN, and ammonia concentrations from samples from a new vintage (2018) could be accurately predicted by a

calibration model built based on samples from the previous two vintages (2016 and 2017). In other words, 2016 and 2017 grape juice samples were used as the calibration set to train the model, while 2018 was used as an independent test set. Furthermore, it was tested to see how accurately the nitrogen status of red grape varieties could be predicted based on a calibration model built from white grape varieties, and vice versa. These calibration models included samples from all three vintages, origins, and the respective red or white varieties.

4.3. Results and Discussion

4.3.1. Tasks and rationale

The rationale of the 66/34 global model was to test the viability and subsequently, the robustness of IR spectroscopy in an industrial context – where samples originate from different varieties, growing regions and vintages. Nicolaï *et al*. (2007) considered a calibration model robust when the model could accurately predict the property of interest, irrespective of unknown changes occurring in the external environment. Due to the innate complexity of fruits and vegetables, samples belonging to different 'batches' (*i.e.* different varieties, origins and vintages) are considered as the most important factor influencing model robustness in the application of IR spectroscopy to agricultural systems (Wang *et al.*, 1991; Nicolaï *et al.*, 2007). This is an important factor to consider in the field of spectroscopy as an inherent feature of this technology is to look at the matrix in its entirety, and subsequently the interactions occurring in the given matrix (Cozzolino *et al.*, 2009). Furthermore, robustness was ensured by assessing models with an independent validation set which avoids potentially overoptimistic results that could be obtained by using a cross-validation strategy.

A subsequent model was built with the calibration and independent validation set adjusted to a ratio of 50/50. This was done to further assess the robustness of the models built by a particular instrument as less samples are included in the training set, as well as increasing the number of independent samples the model is required to predict.

Furthermore, due to the number of environmental factors that influence the grapevine during the growing season, a specific vine may result in a substantially different grape juice matrix from one year to the next. This is known as the vintage effect (Young *et al.*, 2016). Practically speaking, a calibration model would be built using samples from previous vintages and then used to predict the concentration values of samples from a new vintage. Therefore, the next task assigned to each instrument included building a calibration model from two vintages (2016 and 2017) and using it to independently predict the samples from a new vintage (2018). Again, to ensure a realistic situation and increase the robustness, the samples from all the vintages (including both calibration and validation sets) included samples from an array of different cultivars and growing conditions.

The final task given to the instruments were to see how accurately calibrations built using white grape varieties could predict the nitrogen status of red grape varieties and vice versa. This was done due to differences that may occur in the white grape juice matrix compared to red because of the genetics of the grapevine (Chapter 3) as well as the difference in the processing of the grapes during the winemaking process.

4.3.2. Nitrogen status of samples

A total of 905 samples were scanned on each instrument and used for the calibration and validation of the global models. These samples had reference concentrations which spanned over a range of 44.88-483.67 mg N/L, 29.83-365 mg N/L and 1.16-344.97 mg N/L for YAN, FAN and ammonia, respectively (Table 4.1). These concentrations are comparable to what has previously been published for various YAN surveys in other wine regions of the world (Butzke, 1998; Nicolini *et al.*, 2004; Hagen *et al.*, 2008). Thus, another concern of spectroscopic calibration was addressed by ensuring that a large number of samples were collected over a realistic range of concentration values. This dataset is therefore regarded as representative and thus most likely capable of robust calibration of IR spectroscopic instruments for the accurate prediction of the nitrogen status of the grape juice matrix.

The dataset used to test the ability of predicting the nitrogen status of a sample from a new vintage had 799 samples included in the calibration set (2016 and 2017) and 106 in the validation set (2018). The ranges (and other descriptive statistics) of the validation and calibration set can be seen in table 4.2. Moreover, the dataset used to test the ability of the various spectroscopic instruments to test the nitrogen status of red grape juice samples based on white, and vice versa contained 642 white samples and 261 red samples. The summary statistics of this dataset can be seen in table 4.3.

Table 4.1. Descriptive statistics of the reference concentrations used for the calibration and validation of the global models (66/34 and 50/50).

	Global Model		
	YAN	**FAN**	**Ammonia**
Standard deviation	65.21	46.14	24.72
Average	189.61	136.41	53.20
Coefficient of variance	44.69	44.69	44.69
Min	44.88	29.83	1.16
Lower quartile	143.10	104.13	34.22
Median	181.21	130.08	51.77
Upper Quartile	231.95	164.22	69.02
Max	483.67	365.00	167.11
Range	438.79	335.17	165.95

Table 4.2. Descriptive statistics of the reference concentrations used for the calibration and validation of the models used to predict the nitrogen status of a new vintage

	2016 + 2017			2018		
	YAN	**FAN**	**Ammonia**	**YAN**	**FAN**	**Ammonia**
Standard deviation	65.24	46.14	38.46	73.82	47.03	31.52
Average	189.31	134.94	60.52	182.05	124.74	57.30
Coefficient of variance	34.46	34.19	63.55	40.55	37.70	55.01
Min	44.88	29.83	1.16	59.09	44.31	8.68
Lower quartile	142.71	99.44	34.13	131.19	88.96	33.87
Median	180.93	126.57	52.84	169.22	119.76	49.84
Upper Quartile	231.60	163.68	71.68	214.35	142.38	79.59
Max	483.67	365.00	344.97	388.04	269.30	147.64
Range	438.79	335.17	343.81	328.95	224.99	138.96

Table 4.3. Descriptive statistics of the reference concentrations used for the calibration and validation of the models used to predict the nitrogen status of white cultivars based on red, and vice versa.

	White			Red		
	YAN	**FAN**	**Ammonia**	**YAN**	**FAN**	**Ammonia**
Standard deviation	58.71	42.82	21.44	74.29	50.18	30.19
Average	198.69	142.60	56.09	166.25	120.52	45.73
Coefficient of variance	29.55	30.03	38.22	44.69	41.63	66.01
Min	44.88	32.28	8.56	54.34	29.83	1.16
Lower quartile	158.99	113.49	40.78	117.43	87.65	23.56
Median	193.30	138.42	55.33	147.48	106.95	37.79
Upper Quartile	239.05	169.19	70.16	190.81	136.85	58.38
Max	469.41	341.00	167.11	483.67	365.00	147.64
Range	424.53	308.73	158.55	429.33	335.17	146.48

4.3.3. Assessment of IR spectroscopy for the purpose of nitrogen status quantification

Each IR spectroscopy instrument will be discussed individually with regards to firstly, performance in the global tasks (66/34 and 50/50), and subsequently, performance in tasks assigned to assess robustness (*i.e.* predicting the nitrogen status of a new vintage based on previous vintages or of red cultivars based on white, and vice versa). Models will primarily be compared in the discussion based on their RPD and RMSEP values.

4.3.3.1. Fourier-transform infrared spectroscopy prediction models

Strong water absorption peaks (1552-1755 cm^{-1}; 3552-3042 cm^{-1}) can be observed in the FT-IR spectra. This characteristic of FT-IR spectroscopy has been reported to impede its use in quantification of various compositional parameters in the grape juice matrix (Ricci *et al.*, 2014). For example, the peak ranging between 1552-1755 cm^{-1} coincides with the absorption of the amino acid side-chains which absorbs between 1480-1800 cm^{-1} (Barth, 2000). Furthermore, sugar and water absorbing at 3552-3042 cm^{-1} overlap with the 1° N-H_2 groups present in YAN. However, all the models built using FT-IR spectroscopy in transmission mode produced models suitable for quantification as all RPD_{VAL} values were observed to be above 3.

Generally, the global models for all the parameters (YAN, FAN, and ammonia) – for both tested ratios (66/34 and 50/50) (Table 4.4 and 4.5) – were found to perform better than the tasks of predicting the nitrogen status of samples from a new vintage (Table 4.6) or of different colour (red or white) (Table 4.7). Furthermore, global models employing the 66/34 ratio performed better than the 50/50 ratio. RPD_{VAL} values of the 66/34 approach were all found to be above 4 with a RPD_{VAL} of 5.2 obtained for the prediction of total YAN – considered appropriate for quality control purposes (Shah *et al.*, 2010). The 66/34 ratio was found to have the lowest error in prediction for all parameters tested as a RMSEP of 13.9, 11.8 and 5.07 mg N/L was observed for YAN, FAN and ammonia, respectively. The models built based on the 50/50 ratio of calibration/validation were, however, comparable to the models employing the 66/34 ratio as RPD_{VAL} values were also generally observed to be above 4, except for FAN (RPD_{VAL} 3.89). For both ratios, the prediction of FAN was found to be a more difficult task, resulting in a lower RPD_{VAL} compared to YAN and ammonia. Interestingly, although a decrease in the RPD_{VAL} was observed for FAN for the 50/50 ratio compared to the 66/34 ratio, a slight improvement in the average prediction accuracy could be observed for the 50/50 global model (Table 4.4 and 4.5). Furthermore, the rank for the global models (66/34 and 50/50) were observed to range between 16 and 20 (Table 4.4 and 4.5).

A SEP of 5.9 mg N/L and an RPD of 7.8 was obtained by Skoutelas et al. (2011) for the calibration of YAN using FT-IR. The higher RPD and lower error of prediction obtained in this study is most likely due to the model only receiving samples from a single vintage (n=71), the removal of a large number of samples considered to be outliers (n=28/71), as well as the model not undergoing any external validation.

The models built to predict a new vintage also performed accurately, with RPD_{VAL} and rank values of 4.24 and 13, 3.84 and 17, and 4.23 and 18 for YAN, FAN and ammonia, respectively (Table 4.6). Furthermore, the error in prediction obtained by this model (17.6, 11.5 and 7.32 mg N/L, for YAN, FAN and ammonia, respectively) was comparable to what was observed for both global models. Therefore, using FT-IR spectroscopy to predict the nitrogen status of grape juice samples (from an

array of varieties and origins) from a new vintage has proven to be a viable possibility. Thus, by testing the robustness of the models by adding samples from a different growing season, this study has managed to successfully address one of the major concerns regarding the application of this technology in agriculture. However, it must be kept in mind that these calibrations still need to be updated and maintained in the future to ensure that the accuracy and robustness is maintained (Dambergs *et al.*, 2015).

The prediction of red varieties from white, and vice versa proved to be the most difficult task; however, FT-IR still produced good results, with RPD_{VAL} of 3.77 (YAN), 3.24 (FAN) and 4.33 (ammonia) for the prediction of red varieties from a calibration model based on white varieties and RPD_{VAL} of 3.02 (YAN), 3.17 (FAN) and 3.34 (ammonia) for the prediction of white varieties from red and rank values were observed to range between 9 and 20 (Table 4.7). Thus, in this study, the models using white varieties as a calibration set for FT-IR spectroscopy performed better. This is most likely due to the greater variability (range of YAN, FAN, and ammonia concentrations) as well as the larger number of samples of white grapes than red. Therefore, these results emphasise the importance of having a representative dataset for spectroscopic calibration as well as illustrating the need for variability in the dataset. Furthermore, although the RMSEP was generally found to be higher than for the 66/34, 50/50, and vintage models, the RMSEP for both of these tasks (red vs. white and vice versa) were still found to be within an acceptable range (Table 4.7).

Table 4.4. Summary statistics of the global models with calibration/validation ratio of 66/34.

Global Model: Calibration/Validation: 66/34											
Column1	Column2	**N**	**Range (mg/L)**	**Pre-processing**	**Rank**	**R^2_{CAL}**	**RMSEC (mg/L)**	**RPD_{CAL}**	**R^2_{VAL}**	**RMSEP (mg/L)**	**RPD_{VAL}**
FT-IR	YAN	893	53.27-470.5	None	20	94.56	14.5	4.29	96.25	13.9	5.2
	FAN	882	32.28-342.9	First Derivative	16	92.67	11.9	3.69	94.03	11.8	4.09
	Ammonia	886	6.63-167.1	First Derivative	20	95.79	4.95	4.87	95.32	5.07	4.63
FT-NIR	YAN	889	53.27-470.5	None	18	95.06	14	4.5	95.77	14.5	4.87
	FAN	887	32.28-342.9	None	18	91.01	12.7	3.33	91.47	14.5	3.43
	Ammonia	887	8.64-127.6	Constant Offset Elimination	20	90.18	7.62	3.19	87.94	8.47	2.9
ATR-MIR	YAN	885	63.08-438.1	None	15	87.19	22.4	2.79	82.22	24.8	2.07
	FAN	879	32.28-267.1	Constant Offset Elimination	11	79.41	19	2.2	76.23	22.7	2.05
	Ammonia	871	6.09-127.6	Constant Offset Elimination	14	74.54	10.7	1.98	71.71	13.2	1.88

Table 4.5. Summary statistics of the global models with calibration/validation ratio of 50/50.

Global Model: Calibration/Validation: 50/50											
Column1	Column2	**N**	**Range (mg/L)**	**Pre-processing**	**Rank**	**R^2_{CAL}**	**RMSEC (mg/L)**	**RPD_{CAL}**	**R^2_{VAL}**	**RMSEP (mg/L)**	**RPD_{VAL}**
FT-IR	YAN	886	44.8-469.4	First Derivative	18	94.25	15.6	4.17	94.3	15.4	4.19
	FAN	883	32.28-342.9	First Derivative	19	94.09	11.6	4.11	93.18	11.5	3.89
	Ammonia	886	1.16-167.1	None	20	95.87	4.84	4.92	94.45	5.77	4.25
FT-NIR	YAN	891	53.27-470.5	None	17	95.63	14.1	4.78	94	15.6	4.09
	FAN	887	32.28-342.9	None	18	92.96	12.8	3.77	89.15	14.7	3.08
	Ammonia	883	1.16-167.1	None	20	90.23	7.61	3.2	86.43	9.12	2.72
ATR-MIR	YAN	879	53.27-438.1	Constant Offset Elimination	15	87.33	23.5	2.81	81.06	26.9	2.30
	FAN	877	32.28-267.1	Constant Offset Elimination	13	84.61	17.8	2.55	75.13	21.1	2.01
	Ammonia	879	6.09-127.6	None	13	75.2	11.8	2.01	66.55	13.1	1.73

Table 4.6. Summary statistics of the models built to predict the nitrogen status of a new vintage.

Column1 **Vintage Model: Calibration/Validation: 2016+2017/2018** Column2											
		N	**Range (mg/L)**	**Pre-processing**	**Rank**	**R^2_{CAL}**	**RMSEC (mg/L)**	**RPD_{CAL}**	**R^2_{VAL}**	**RMSEP (mg/L)**	**RPD_{VAL}**
FT-IR	YAN	893	59.09-388	MSC	13	91.75	18.5	3.48	94.36	17.6	4.24
	FAN	882	44.31-267.9	None	17	91.74	12.7	3.48	93.11	11.5	3.84
	Ammonia	886	8.68-147.6	MSC	18	93.83	5.77	4.03	94.4	7.32	4.23
FT-NIR	YAN	892	59.09-388	None	16	93.8	16	4.02	94.36	17.5	4.26
	FAN	888	44.31-269.3	Constant Offset Elimination	17	89.64	14.3	3.11	91.49	13.7	3.43
	Ammonia	882	8.68-135.6	Constant Offset Elimination	20	89.33	7.66	3.06	91.83	8.46	3.51
ATR-MIR	YAN	892	59.09-388	SNV	8	70.23	32.9	1.83	75.83	34.1	2.05
	FAN	883	44.31-267.9	Min-Max Normalization	11	75.26	21.2	2.01	77.46	21.3	2.17
	Ammonia	875	8.68-120.2	First Derivative + SNV	7	63.86	13.2	1.66	61.62	16.7	1.62

Table 4.7. Summary statistics of the models built to predict the nitrogen status of white cultivars based on white, and vice versa.

Red vs. White													
Column1	Column2	**Calibration**	**Validation**	**N**	**Range (mg/L)**	**Pre-processing**	**Rank**	**R^2_{CAL}**	**RMSEC (mg/L)**	**RPD_{CAL}**	**R^2_{VAL}**	**RMSEP (mg/L)**	**RPD_{VAL}**
FT-IR	**YAN**	white	red	882	54.34-470.5	MSC	20	94.8	13.6	4.38	92.94	18.3	3.77
		red	white	881	44.88-469.4	Constant offset Elimination	16	94.55	18	4.28	88.97	19.5	3.02
	FAN	white	red	881	41.98-342.9	First Derivative	12	92.11	11.7	3.56	90.2	14.6	3.24
		red	white	881	32.28-341	None	9	91.77	14.4	3.49	90.03	13.4	3.17
	Ammonia	white	red	883	1.16-132.7	Constant Offset Elimination	19	95.02	4.85	4.48	94.66	6.2	4.33
		red	white	882	8.56-167.1	None	20	96.93	5.37	5.71	91.02	6.44	3.34
FT-NIR	**YAN**	white	red	875	54.34-388	SNV	13	94.08	13.9	4.11	89.09	22.1	3.08
		red	white	887	44.88-469.4	SNV	14	95.94	15.4	4.96	92.33	16.2	3.61
	FAN	white	red	876	41.98-342.9	First Derivative	18	89.52	13.8	3.09	86.44	17.3	2.72
		red	white	876	32.28-341	None	20	95.29	10.8	4.61	86.43	15.7	2.73
	Ammonia	white	red	881	1.16-127.6	None	20	92.09	5.83	3.56	88.8	9.15	2.99
		red	white	883	8.56-123.8	Constant offset Elimination	19	94.49	7.26	4.26	81.56	8.78	2.33

Table 4.7. (cont.)

Red vs. White													
Column1	Column2	**Calibration**	**Validation**	**N**	**Range (mg/L)**	**Pre-processing**	**Rank**	**R^2_{CAL}**	**RMSEC (mg/L)**	**RPD_{CAL}**	**R^2_{VAL}**	**RMSEP (mg/L)**	**RPD_{VAL}**
ATR-MIR	**YAN**	white	red	874	54.34-483.7	None	16	81.83	24.5	2.35	71.01	35.1	1.86
		red	white	882	53.27-437.7	Straight line subtraction	6	76.06	36.7	2.04	62.2	34.6	1.63
	FAN	white	red	864	41.98-285.6	Constant offset Elimination	18	82.01	17.8	2.36	60.86	27.7	1.73
		red	white	869	32.28-315.5	None	10	84.79	18.5	2.56	62.98	25.1	1.63
	Ammonia	white	red	876	6.09-127.2	SNV	15	77.82	9.82	2.12	70.9	14.1	1.87
		red	white	863	8.56-122.2	Straight line subtraction	6	73.69	13.8	1.95	53.36	14	1.65

4.3.3.2. Fourier-transform near-infrared spectroscopy

The NIR spectra, characterised by the overtones and combination bands caused by the fundamental vibrations occurring in the mid-infrared range, was dominated by the overtones of the O-H stretch (7274-6338 cm^{-1}) and a combination band of O-H stretching and bending (5417-4495 cm^{-1}), induced by the presence of water in the grape juice matrix (Büning-Pfaue, 2003). Despite this, NIR spectroscopy has been reported to be appropriate for quantification purposes as the band shape is often typical of a specific compound or a group of compounds (Ricci *et al.*, 2014).

As with FT-IR spectroscopy, the 66/34 global model performed the best when looking at both the RPD_{VAL} and RMSEP statistics (Table 4.4). A better RPD_{VAL} was, however, observed for the prediction of ammonia concentrations of samples from a new vintage (RPD_{VAL} of 3.51 compared to 2.9 for the 66/34 model) although, the difference between the two models in terms of the RMSEP was considered irrelevant (8.47 vs 8.46 mg N/L for the 66/34 and 2016+2017/2018 model, respectively). Furthermore, RPD_{VAL} value for the 66/34 global model to predict total YAN was also close to 5, as was the case for FT-IR spectroscopy. In terms of the other parameters, higher RPD_{VAL} values were obtained for FAN (RPD_{VAL} 3.43 and 3.08) compared to ammonia (RPD_{VAL} 2.9 and 2.72), for both global models (66/34 and 50/50, respectively) for FT-NIR spectroscopy. This is in contrast to what was found for FT-IR, where ammonia was found to be more accurately predicted than FAN. As RPD_{VAL} values for FT-NIR were found to be more than 3 for YAN and FAN for both global model ratios, this method was found to be adequate for accurate quantification of these parameters (Nicolaï *et al.*, 2007). Although decreased accuracy was obtained for the quantification of ammonia (RPD_{VAL} < 3), these values are still deemed satisfactory (RPD_{VAL} > 2.5) (Nicolaï *et al.*, 2007). Furthermore, the rank of these models was observed to range between 17-20 (Table 4.4 and 4.5).

The task of predicting a new vintage (Table 4.6) resulted in higher RPD_{VAL} values than for the 50/50 global models (Table 4.5). This may be due to the larger number of samples used to train these models in addition to the reduced number of samples tested against these models. Furthermore, this model also outperformed the 50/50 global model in terms of the RMSEP for FAN and ammonia, obtaining errors of 13.7 and 8.46 mg N/L, respectively. Rank of these models ranged between 16-20. Interestingly, the prediction of total YAN of a sample from a new vintage using FT-NIR was observed to be (although marginally), better than what was found for FT-IR spectroscopy (Table 4.6). The results for FT-NIR spectroscopy to predict the FAN and ammonia concentrations of a new vintage were also considered to be adequate for accurate quantification (RPD_{VAL} > 3) (Nicolaï *et al.*, 2007). Therefore, FT-NIR spectroscopy can be considered a viable technique for the prediction of samples from a new vintage and a feasible option for industrial use.

Again, the task of predicting the nitrogen status of a red grape juice sample based on a calibration model including only white cultivars, and vice versa, was generally found to be the most challenging.

This showed in the lower RPD_{VAL} values as well as the higher RMSEP obtained for these tasks (Table 4.7). Rank was observed to range between 13 and 20. The more accurate quantification of YAN compared to FAN and ammonia was also observed for these tasks (with $RPD_{VAL} < 3$ being obtained, compared to $RPD_{VAL} > 3$ obtained for the prediction of YAN). Although, as RPD_{VAL} values for FAN and ammonia were still generally found to be more than 2.5 (except for the prediction of ammonia of red cultivars based on white: RPD_{VAL} 2.33), FT-NIR spectroscopy is also considered to be a reasonably robust method.

4.3.3.3. Attenuated total reflectance mid-infrared spectroscopy

ATR-MIR spectra of the grape juice samples were mainly characterised by a strong sharp peak at 950-1100 cm^{-1}, corresponding to water peaks, whereas peaks occurring between 1480-1800 cm^{-1} are related to C=N, C=C and C=O stretching and N-H bending, corresponding to bonds found in amino acids and their side chains (Barth, 2000). The carboxylic acid O-H stretch produced peaks between 2800-2970 cm^{-1} which can be owed to amino acids as well as organic acids present in the grape juice medium and therefore, this could lead to interferences in the spectra, hampering accurate quantification. Furthermore, the presence of sugars can also interfere with accurate quantification due to the sp^3 C-H stretch found in this region as well as the alcohol O-H stretch occurring between 3388-3600 cm^{-1}, coinciding with primary and secondary amino nitrogen groups (1°N-H_2; 2°N-H_2).

Overall, ATR-MIR was not found to be suitable for accurate quantification purposes as RPD_{VAL} values were never observed to be more than 2.5 (Nicolaï *et al.*, 2007), with many found to be less than 2 (Table 4.4-4.7). Rank values for ATR-MIR were generally lower than for other spectroscopies; ranging between 11-15. However, following the trend of the abovementioned spectroscopies, both global models were still found to be generally more accurate than what was observed for the other tasks. The highest RPD_{VAL} was obtained for the prediction of YAN in the 50/50 global (RPD_{VAL} 2.3), however, a higher RMSEP was obtained for this model (26.9 mg N/L) compared to the 66/34 model (24.8 mg N/L; RPD_{VAL} 2.07). Furthermore, as with FT-NIR spectroscopy, higher RPD_{VAL} were obtained for YAN and FAN ($RPD_{VAL} > 2$) compared to ammonia ($RPD_{VAL} < 2$). This trend was not only observed for the global models, but generally throughout the tasks of robustness assigned to the instrument.

The prediction of samples originating from a new vintage was again observed to be more accurate than to predict red cultivars based on white, and vice versa. However, the prediction of the ammonia concentrations of white cultivars based on red was found to be an exception (RPD_{VAL} 1.87; RMSEP 14.1 mg N/L). RPD_{VAL} for the prediction of a new vintage ranged between 1.62 (ammonia) and 2.17 (FAN). Together with the lower RPD_{VAL}, higher errors in prediction (RMSEP) were observed for this task (Table 4.6) compared to the global models (Table 4.4 and 4.5) as well as compared to the other spectroscopies for the same task. Again, rank values were observed to be lower than for other spectroscopies, ranging between 7-11.

Furthermore, the lowest RPD_{VAL} (ranging between 1.51-1.87) and highest RMSEP found in the study was obtained for the prediction of red cultivars based on white, and vice versa. As such, ATR-MIR spectroscopy was considered to be less robust than FT-IR and FT-NIR spectroscopy for the prediction of the nitrogen status of the grape juice matrix. Rank values ranged between 6-16. The only other study, to our knowledge, testing the viability of using ATR-MIR for the prediction of the nitrogen status of grape juice samples is a study by Shah *et al.* (2010). However, the cited study made use of cross-validation instead of an independent test, obtaining RPD_{CV} of 2 for YAN and FAN and 2.1 for ammonia. These results obtained by Shah *et al*. (2010) are comparable to the results found in the present study, and even though the RMSEP found here is in most cases less than what was found by Shah *et al*. (2010), the RPD in the validation step found in our study still did not improve dramatically, if at all. Therefore, ATR-MIR may be able to distinguish high values from low and thus, may be suitable for screening purposes. However, this spectroscopy technique is not considered sufficient for accurate quantification of grape juice nitrogen status to help ensure optimal fermentation conditions.

4.3.4. Overall Trends

4.3.4.1. Comparison of the performance of the instruments

Overall, for each instrument, total YAN predictions were observed to be more accurate than measuring the components separately. This was shown through the higher RPD values obtained for YAN than for FAN and ammonia separately, as well as the lower error in prediction (RMSEP) found for YAN compared to the sum of the errors obtained for FAN and ammonia (Tables 4.4-4.7). Furthermore, for all tasks (global, vintage and red vs. white models) FT-IR was able to predict total YAN and ammonia more effectively than FAN, whereas FT-NIR and ATR-MIR was able to predict total YAN and FAN more effectively than ammonia.

Taken together, FT-IR (WineScan™ FT120) outperformed both other instruments for the measurement of all three of the investigated parameters, throughout all the given tasks. This is because consistently higher RPD_{VAL} as well as lower RMSEP were observed for this instrument compared to the other spectroscopies. However, the MPA, measuring in the NIR range in transmission mode, also produced models capable of accurate quantification, although the validation statistics were slightly less optimal than what was found for FT-IR. It would, however, be advisable to rather use FT-IR for the quantification of ammonia compared to FT-NIR as FT-IR obtained $RPD_{VAL} > 4$ compared to < 3 for FT-NIR.

ATR-MIR was, however, not comparable to either FT-IR or FT-NIR spectroscopy for any of the parameters or tasks assigned. This is due to the consistently lower RPD_{VAL} and higher RMSEP obtained throughout. Thus, this instrument is only suitable for screening purposes and not for the accurate quantification of any of the parameters tested. It was surprising that the FT-NIR spectral

instrument outperformed the ATR FT-MIR instrument as MIR spectra are produced due to the fundamental stretching, bending and rotating vibrations produced by various functional groups present in the sample. On the other hand, spectral signatures in the near infrared region are only due to the complex overtones of these fundamental vibrations. Furthermore, the combination bands, such as those produced by C-O stretch and the N-H band in protein, as well as water, which is a major component of most fruits and vegetables, can result in a highly convoluted NIR spectrum, decreasing the chances of accurate quantification and interpretation (Cozzolino, 2015; Nicolaï *et al*., 2007). However, the regions that were selected for the optimization of the models for the FT-IR models primarily fell within the mid-infrared range (YAN ~4200-1200 cm^{-1}; FAN ~4600-1400 cm^{-1}; ammonia ~3000-1200 cm^{-1}). Thus, it is hypothesised that the mode that the spectra was collected in (reflectance vs. transmission) also played a major role in the difference in performance obtained between the instruments and thus, transmission was found to be more suitable than reflectance for this application.

4.3.4.2. Trends in pre-processing techniques applied

Pre-processing is a useful technique to enhance the quality of the calibration and therefore, increase the chances of obtaining an accurate prediction. This is achieved through the removal of any irregularities found in the spectra that are due to non-constituent interferences. These interferences include scattering, shifts in the baseline or wavelength and noise induced by the detector, for example (Yahia, 2017).

For FT-IR spectroscopy, for the global tasks, pre-processing techniques resulting in the most accurate models included first derivative or no pre-processing, depending on the parameter to be predicted (YAN, FAN, or ammonia) as well as the ratio of the calibration to the validation set (66/34 and 50/50). First derivative pre-processing has been reported as an appropriate method of pre-processing FT-IR spectral data as it helps to decrease the baseline shift and avoids the intensity effect, which are typical characteristics of FT-IR spectra (Yahia, 2017). No spectral pre-processing was required for the prediction of FAN concentrations from a new vintage (*i.e.* where 2016 and 2017 calibration models were used to predict independent samples from 2018). However, for this task, multiplicative scattering correction (MSC) enhanced the accuracy of the models for the prediction of YAN and ammonia concentrations. This form of pre-processing is one of the most widely used methods and is used to compensate for the baseline shift in addition to accounting for multiplicative effects triggered by non-uniform scattering of IR light (Rinnan *et al.*, 2009).

No particular trend could be observed for the pre-processing technique that resulted in the most accurate models to predict the nitrogen status of white varieties from red, or vice versa for FT-IR spectroscopy. The techniques used included a combination of first derivative, MSC, as well as constant offset elimination; however, sometimes no pre-processing was required.

Constant offset elimination was, however, a technique frequently resulting in the most accurate calibrations for both FT-NIR spectroscopy as well as ATR-MIR. This technique helps to correct baseline shifts as well as enhancing the innate absorption properties of the grape juice matrix (Yahia, 2017). As such, more relevant information can be extracted from the original spectra, allowing for more accurate predictions. This technique, together with no spectral pre-processing was used to optimize the global models (34/66 and 50/50) for both aforementioned spectroscopies, as well as predicting the nitrogen status of a new vintage for FT-NIR spectroscopy and the prediction of white varieties based on red for ATR-MIR spectroscopy. The other pre-processing technique frequently used to optimise calibration models for FT-NIR and ATR-MIR spectroscopy was standard normal vector (SNV), which also corrects baseline shifts as well as scattering effects that occur due to path length differences (Wang *et al.*, 2006) . However, overall, the majority of FT-NIR prediction models did not require any pre-processing to produce accurate and robust models.

Techniques such as straight-line subtraction and a combination of first derivative and SNV were also applied to ATR-MIR spectroscopy. Applying first derivative followed by SNV, effectively perform the same task as straight-line subtraction; correcting for the baseline shift in the same way (Yahia, 2017).

4.3.5. YAN, FAN and ammonia in context

In order to assess whether the models produced in this study are accurate enough for industrial use, it is important to understand the parameters, YAN, FAN and ammonia, in the context of the winemaking environment. Yeast assimilable nitrogen is an essential nutrient required by yeast during fermentation. In the absence of sufficient concentrations, yeast will not be able to produce the required amounts of biomass that is necessary to carry a fermentation through to dryness, and therefore, fermentations may become stuck or sluggish (Henschke & Jiranek, 1993; Bisson, 1999). In addition to the large amounts of residual sugar that will be present in the wine, stuck or sluggish fermentations are normally accompanied by the formation of off-flavours, such as H_2S (Gobbi *et al.*, 2013). Furthermore, insufficient concentrations of the FAN component of YAN has been reported to lead to a very neutral wine devoid of desirable fruity and floral aromas. This is because the branched-chain and aromatic amino acids (which form part of the FAN component of YAN) have been identified as the precursor molecules for the formation of these favourable aromas (Rapp & Versini, 1991; Smit, 2013).

The exact amount of YAN, FAN and ammonia which is optimal for the yeast during fermentation is highly strain dependent, however, a 140 mg N/L of total YAN has been benchmarked in literature as the minimum amount required to complete fermentation (Bely *et al.*, 1990). The range of YAN, FAN and ammonia concentrations found in various surveys across different wine regions can be can be found in Chapter 3. Studies done to investigate the impact of varying concentrations of YAN, FAN and ammonia on the fermentation efficiency and organoleptic qualities of the final wine have found that, at above a certain threshold, the amount of YAN becomes redundant. For example, the

production of fruity and floral esters has been observed to plateau when total YAN concentrations reach more than 250-300 mg N/L, and have even been found to decrease when YAN concentrations reach approximately 500 mg N/L (Vilanova *et al.*, 2007). Furthermore, very high total YAN concentrations (>450-500 mg N/L) may result in the production of unwanted compounds such as biogenic amines, carcinogens and protein haze, as well as leading to microbial instability (Bell & Henschke, 2005). Therefore, having excessive concentrations of YAN will decrease the quality of the final product.

These margins of concern are, however, over approximately a 50 mg N/L (total YAN) range, depending on the nitrogen demand of the particular yeast strain used. Therefore, the use of ATR-MIR may be plausible from a screening point of view but will not allow for precise decision-making regarding nitrogen supplementation. It is important to note that the RMSEP reported is an *average* of the errors and that, in some cases, this error may be a lot larger than the value reported as the RMSEP. Therefore, there is a chance that winemakers may be completely misguided by the prediction value given by ATR-MIR.

Due to the different roles that the components of YAN (FAN and ammonia) play in the metabolic activities of the yeast, as well as the high costs involved in nutrient supplementation, having a more precise indication of the nitrogen status becomes essential. This is in addition to the rise in studies investigating the possibility of modulating wine organoleptic qualities by manipulation of the nitrogen status of the must to either suit a particular strain of yeast, or a particular must composition (Ugliano *et al.*, 2007; Vilanova *et al.*, 2007; Garde-Cerdán *et al.*, 2011; Barbosa *et al.*, 2012; Rollero *et al.*, 2018). Therefore, having an accurate method of quantification will allow for more fine-tuned and informed decision-making with regards to nutrient supplementation, providing winemakers with the opportunity to optimize the fermentation for the desired style and quality.

In light of this, using FT-IR, or even FT-NIR spectroscopy would be more beneficial than ATR-MIR as there are lower RMSEP and higher RPD_{VAL} values. High RPD values are important as the RPD of a model is an indicator of how reliable the model is i.e. it indicates how reliable the RMSEP of the model is. Furthermore, the RMSEP reported for these two instruments are low enough in the context of the YAN status of grape must to allow for optimal and precise nitrogen supplementation.

4.4. Conclusion

To the authors' knowledge, this is the first study of its kind, incorporating such a large degree of variability for the purpose of quantifying the nitrogen status of the grape juice matrix. This variability is demonstrated by the large number of samples as well as the number of different grape varieties, origins, and vintages incorporated in both the calibration and validation sets. In addition to this, an

independent validation set was used. This is a shortcoming highlighted in most other studies in this field which impedes the widespread use of this technology for routine analysis of fruits and vegetables.

The results obtained in this study show that it is indeed possible to calibrate IR spectroscopic instruments for the accurate measurement of YAN, FAN, and ammonia concentrations. Transmission FT-IR spectroscopy was, however, observed to show the most promising results; however, FT-NIR spectroscopy also produced models capable of good to excellent quantification, primarily for YAN and FAN. Furthermore, both of these instruments showed sufficient robustness against samples originating from different varieties, growing conditions, and vintages, addressing the concerns of applying this technology to the agricultural industry. Therefore, applying this rapid, cost-effective, and environmentally friendly method in an industrial setup is a plausible option, despite the inherent variability and complexity of the grape juice matrix. Moreover, the possibility of measuring the YAN status of samples from a new vintage are one of the most important findings in this study as it demonstrates the feasibility of this technology in an industrial set-up. This is because calibrations will most likely be based on samples originating from previous vintages and used for analysis of subsequent vintages.

However, the importance of obtaining a representative calibration set is highlighted through the diminished performance of models used to predict red varieties from white, and vice versa. Therefore, the differences between red and white grape varieties are substantial enough to have an impact on the prediction ability of the models and samples from each category should be included in the calibration set to ensure better predictions.

Due to the accuracy, robustness, high throughput, and cost-effective nature, the models produced by both FT-IR and FT-NIR spectroscopy provide winemakers with the opportunity to make more timely and informed nutrient supplementation decisions, facilitating the achievement of their desired wine style and quality.

References

Aleixandre-Tudo, J.L., Nieuwoudt, H., Aleixandre, J.L., du Toit, W., 2018. Chemometric compositional analysis of phenolic compounds in fermenting samples and wines using different infrared spectroscopy techniques Talanta 176, 526–536.

Anderssen, E., Dyrstad, K., Westad, F., Martens, H., 2006. Reducing over-optimism in variable selection by cross-model validation. Chemometrics and Intelligent Laboratory Systems 84, 69–74.

Barbosa, C., Mendes-Faia, A., Mendes-Ferreira, A., 2012. The nitrogen source impacts major volatile compounds released by Saccharomyces cerevisiae during alcoholic fermentation. International Journal of Food Microbiology 160(2), 87–93.

Barth, A., 2000. The infrared absorption of amino acid side chains. Progress in Biophysics and Molecular Biology 74(3), 141–173.

Bauer, R., Nieuwoudt, H., Bauer, F.F., Kossmann, J., Koch, K.R., Esbensen, K.H., 2008. FTIR Spectroscopy for Grape and Wine Analysis. Analytical Chemistry 80(5), 1371–1379.

Bell, S.-J. & Henschke, P.A., 2005. Implications of nitrogen nutrition for grapes, fermentation and wine. Australian Journal of Grape and Wine Research 11(3), 242–295.

Bely, M., Sablayrolles, J.-M., Barre, P., 1990. Automatic detection of assimilable nitrogen deficiencies during alcoholic fermentation in oenological conditions. Journal of Fermentation and Bioengineering 70(4), 246–252.

Bisson, L.F., 1999. Stuck and Sluggish Fermentations. Am J Enol Vitic. 50(1), 107-119.

Büning-Pfaue, H., 2003. Analysis of water in food by near infrared spectroscopy. Food Chemistry 82(1), 107–115.

Butzke, C.E., 1998. Survey of Yeast Assimilable Nitrogen Status in Musts from California, Oregon, and Washington. Am. J. Enol. Vitic., 49(2), 220-224.

Cozzolino, D., 2009. Near Infrared Spectroscopy in Natural Products. Analysis Planta Medica 75(7), 746–756.

Cozzolino, D., 2015. The Role of Visible and Infrared Spectroscopy Combined with Chemometrics to Measure Phenolic Compounds in Grape and Wine Samples. Molecules 20, 726–737.

Cozzolino, D., Cynkar, W.U., Shah, N., Dambergs, R.G., Smith, P.A., 2009. A brief introduction to multivariate methods in grape and wine analysis. International Journal of Wine Research 1, 123-130.

Cozzolino, D., Cynkar, W., Shah, N., Smith, P., 2011. Technical solutions for analysis of grape juice, must, and wine: the role of infrared spectroscopy and chemometrics. Anal Bioanal Chem 401(5), 1475–1484.

Dambergs, R.G., Kambouris, B., Cnykar, W.U., Janik, L.J., Cozzolino, D., Henschke, P.A., Gishen, M., 2004. A comparison of near infrared and mid-infrared spectroscopy for thecanalysis of yeast assimilable nitrogen in grape juice. In: Australian Wine Industry Technical Conference Inc: Melbourne, Australia. pp 334–335.

Dambergs, R.G., Gishen, M., et al., 2015. A Review of the State of the Art, Limitations, and Perspectives of Infrared Spectroscopy for the Analysis of Wine Grapes, Must, and Grapevine Tissue Applied Spectroscopy Reviews 50, 3, 261–278.

Garde-Cerdán, T., Martínez-Gil, A.M., Lorenzo, C., Lara, J.F., Pardo, F., 2011. Implications of nitrogen compounds during alcoholic fermentation from some grape varieties at different maturation stages and cultivation systems. Food Chemistry 124(1), 106–116.

Gishen, M., Cozzolino, D., Dambergs, R., G., 2010. The Analysis of Grapes, Wine, and Other Alcoholic Beverages by Infrared Spectroscopy. In: Handbook of Vibrational Spectroscopy, John Wiley & Sons, Ltd. Chichester, UK. pp 1-18.

Gobbi, M., Comitini, F., D'Ignazi, G., Ciani, M., 2013. Effects of nutrient supplementation on fermentation kinetics, H2S evolution, and aroma profile in Verdicchio DOC wine production. European Food Research and Technology 236(1), 145–154.

Gump, B.H., Zoecklein, B.W., Fugelsang, K.C., Whiton, R.S., 2002. Comparison of Analytical Methods for Prediction of Pre-fermentation Nutritional Status of Grape Juice. Am. J. Enol. Vitic. 53(4), 325–329.

Hagen, K.M., Keller, M., Edwards, C.G., 2008. Survey of Biotin, Pantothenic acid, and Assimilable Nitrogen in Winegrapes from the Pacific Northwest. Am. J. Enol. Vitic. 59(4), 432–436.

Henschke, P.A. & Jiranek, V., 1993. Yeasts - metabolism of nitrogen compounds. In: Wine Microbiology and Biotechnology. Harwood Academic Publishers, Chur, Switzerland. pp. 77–164.

Larrain, M., Guesalaga, A.R., Agosin, E., 2008. A Multipurpose Portable Instrument for Determining Ripeness in Wine Grapes Using NIR Spectroscopy. IEEE Transactions on Instrumentation and Measurement 57(2), 294–302.

Liu, F., He, Y., Wang, L., Sun, G., 2011. Detection of Organic Acids and pH of Fruit Vinegars Using Near-Infrared Spectroscopy and Multivariate Calibration. Food and Bioprocess Technology 4(8), 1331–1340.

Manley, M., van Zyl, A., Wolf, E.E.H., 2001. The Evaluation of the Applicability of Fourier Transform Near-Infrared (FT-NIR) Spectroscopy in the Measurement of Analytical Parameters in Must and Wine. South African Journal of Enology & Viticulture 22(2), 93-100.

Martelo-Vidal, M.J. & Vázquez, M., 2014. Evaluation of ultraviolet, visible, and near infrared spectroscopy for the analysis of wine compounds. Czech Journal of Food Sciences 32(1), 37–47.

Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007. Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology 46(2), 99–118.

Nicolini, G., Larcher, R., Versini, G., 2004. Status of yeast assimilable nitrogen in Italian grape musts and effects of variety, ripening and vintage. Vitis 43(2), 89–96.

Patz, C.D., Blieke, A., Ristow, R., Dietrich, H., 2004. Application of FT-MIR spectrometry in wine analysis. Analytica Chimica Acta 513(1), 81–89.

Rapp, A. & Versini, G., 1991. Influence of nitrogen compounds in grapes on aroma compounds of wines In: Developments in Food Science 37. Elsevier 1659–1694.

Ricci, A., Paripinello, G.P., Laghi, L., Lambri, M., Versari, A., 2014. Chapter 2: Application of Infrared Spectroscopy to Grape and Wine Analysis In: Infrared Spectroscopy. Nova Science Publishers, Inc. pp 17-42.

Rinnan, Å., van Den Berg, F., Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. Trends in Analytical Chemistry 28(10), 1201–1222.

Rollero, S., Bloem, A., Ortiz-Julien, A., Camarasa, C., Divol, B., 2018. Fermentation performances and aroma production of non-conventional wine yeasts are influenced by nitrogen preferences. FEMS Yeast Research 18(5), 1-11.

Schaare, P.N., McGlone, V.A., Oliver, R.J. Clark, C.J., 2012. Using Visible/Near Infrared Spectroscopy to Assess Soluble Solids Content of Grapes on a Moving Conveyor In: American Society of Agricultural and Biological Engineers.

Shah, N., Cynkar, W., Smith, P., Cozzolino, D., 2010. Use of Attenuated Total Reflectance Midinfrared for Rapid and Real-Time Analysis of Compositional Parameters in Commercial White Grape Juice. Journal of Agricultural and Food Chemistry 58(6), 3279–3283.

Skoutelas, D., Ricardo-da-Silva, J.M., Laureano, O., 2011. Validation and Comparison of Formol and FT-IR Methods for Assimilable Nitrogen in Vine Grapes. South African Journal of Enology and Viticulture 32(2), 262–266.

Smit, A.Y., 2013. The impact of nutrients on aroma and flavour production during wine fermentation. Thesis, Stellenbosch University, Private Bag X1, 7602 Matieland (Stellenbosch), South Africa.

Ugliano, M., Henschke, P.A., Herderich, M.J., Pretorius, I.S., 2007. Nitrogen management is critical for wine flavour and style. Wine Industry Journal. 22(6), 24-30.

Versari, A., Parpinello, G.P., Mattioli, A.U., Galassi, S., 2008. Determination of Grape Quality at Harvest Using Fourier-Transform Mid-Infrared Spectroscopy and Multivariate Analysis. Am. J. Enol. Vitic. 59(3).

Vilanova, M., Ugliano, M., Varela, C., Siebert, T., Pretorius, I.S., Henschke, P.A., 2007. Assimilable nitrogen utilisation and production of volatile and non-volatile compounds in chemically defined medium by Saccharomyces cerevisiae wine yeasts. Applied Microbiology and Biotechnology 77(1), 145–157.

Wang, L., Lee, F.S.C., Wang, X. He, Y., 2006. Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance NIR. Food Chemistry 95(3) 529–536.

Wang, Y., Veltkamp, D.J., Kowalski, B.R., 1991. Multivariate instrument standardization. Analytical Chemistry 63(23), 2750–2756.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58(2), 109–130.

Yahia, E.M., 2017. Fruit and Vegetable Phytochemicals: Chemistry and Human Health, 2. John Wiley & Sons, Chichester, UK.

Young, P.R., Eyeghe-Bickong, H.A., du Plessis, K., Vivier, M., et al., 2016. Grapevine Plasticity in Response to an Altered Microclimate: Sauvignon Blanc Modulates Specific Metabolites in Response to Increased Berry Exposure. Plant Physiol 170(3), 1235–1254.

Chapter 5

Research Results

Grape Must Profiling and Cultivar Discrimination Based on Amino Acid Composition

Chapter 5
Grape Must profiling and Cultivar Discrimination Based on Amino acid Composition

5.1. Introduction

The grape juice matrix presents the yeast with a complex mixture of nutrients during fermentation. Nitrogenous compounds are, however, one of the most important classes, second only to carbon (Bely *et al.*, 1990a; Stines *et al.*, 2000). As sugar is in most cases present in sufficient quantities to support the growth of the yeast during fermentation, the nitrogen concentration of the must has been identified as the most common cause for stuck or sluggish fermentations (Bisson, 1999). The assimilable portion of the grape juice matrix is referred to as yeast assimilable nitrogen (YAN) and is primarily made up of ammonium and free amino nitrogen (FAN) (Bell & Henschke, 2005).

The free amino nitrogen portion is comprised of an array of amino acids and has been reported to make up 50-90% of the grape must YAN (Kliewer, 1969, 1970). Due to the number of amino acids contributing to FAN, the specific profile and concentration of amino acids has become an important area of research. This is mainly attributed to the complex role that amino acids play in the metabolic activities of the yeast and, subsequently, the effect that it has on the quality of the final wine (Ugliano *et al.*, 2007). This complexity can be illustrated by the fact that not all amino acids are equally substantial in supporting the growth of the yeast, and thus, there is a preferential uptake of certain amino acids (Beltran *et al.*, 2004). Consequently, certain amino acids are denoted as ‘good’ sources of nitrogen and others as ‘poor’. Aside from ammonium, amino acids which are preferred by the yeast include glutamate, glutamine, aspartate, asparagine and arginine, whereas tryptophan, histidine, glycine, and lysine are considered as poor sources of nitrogen (Cooper, 1982; Beltran *et al.*, 2004). On the other hand, proline, the most abundant amino acid (together with arginine) (Ough & Bell, 1980; Stines *et al.*, 2000) is not considered as a source of YAN during fermentative conditions. This is due to the oxygen requirement of the first step involved in proline catabolism (Wang & Brandriss, 1987).

Other than fulfilling the biosynthetic requirement of the yeast, and thereby ensuring optimal fermentation kinetics, the oenological relevance of amino acid metabolism stems from the range of by-products that are subsequently produced. These by-products have been reported to have a significant impact on the organoleptic qualities of the final wine (Rapp & Versini, 1991). Of particular interest is the formation of higher alcohols and esters due to the presence of branched-chain (valine, leucine, and isoleucine) and aromatic (tryptophan, tyrosine, and phenylalanine) amino acids (Rapp & Versini, 1991; Hernández-Orte *et al.*, 2002; Torrea, 2003; Vilanova *et al.*, 2007; Smit, 2013; Rollero *et al.*, 2018). As these amino acids are the precursor molecules for aroma compounds a *direct* link exists between the presence of the amino acid and the corresponding higher alcohol and ester (Rapp

& Versini, 1991). However, a study conducted by Hernández-Orte, Cacho, and Ferreira (2002) found that the amino acid composition influences the concentration of other compounds for which amino acids are *not* the direct precursors. Examples of these include ethanol, acetic acid, as well as fatty acids. This is in support of an earlier report which stated that YAN levels influence all the primary and secondary products of glycolysis which is owed to the involvement of nitrogen in regulating the transport, metabolism and accumulation of sugar by the yeast (Boulton *et al.*, 1999). Furthermore, the nitrogen content of the must has been reported to induce *de novo* synthesis of monoterpenes, previously thought to only originate from the grape berry itself (Carrau *et al.*, 2005).

Additionally, the amino acid content of the must has been found to influence the presence of various unwanted compounds, detrimental to the quality and safety of the wine. A deficiency in the sulphur-containing amino acids, cysteine and methionine, as well as an overall low level of YAN has been linked to the production of hydrogen sulphide (H_2S), known to elicit a rotten egg-like aroma (Swiegers & Pretorius, 2007; Gobbi *et al.*, 2013). Furthermore, arginine has been implicated in the production of a carcinogen, ethyl carbamate, through a spontaneous chemical reaction of ethanol with carbamyl-related compounds such as urea (released by yeast) and citrulline (released by lactic acid bacteria) (Ough *et al.*, 1988; Guo *et al.*, 2016). However, the balance of other amino acids in relation to arginine is also said to play a role (Ough, Crowell & Mooney, 1988; Ough *et al.*, 1991). Moreover, the decarboxylation of amino acids by lactic acid bacteria, typically occurring in conditions of nitrogen excess, has been found lead to the formation of biogenic amines (Smit *et al.*, 2012). As these compounds are known to have potentially harmful physiological effects on human beings, they are a matter of concern for the wine industry (Landete *et al.*, 2007).

The amino acid profile of a particular grape must is a result of a variety of factors. These include the interaction between the genetic background of the vine with the surrounding environment. In other words, an interplay of the grape variety with the climate, soil, and various viticultural practices, exists (Bell & Henschke, 2005; Garde-Cerdán *et al.*, 2009). This knowledge has prompted the investigation of various grape compositional elements in relation to the variety, geographical origin, and vintage of the resulting wine (Soufleros *et al.*, 2003; de Villiers *et al.*, 2005; Camara *et al.*, 2006; Liu *et al.*, 2006; Serrano-Lourido *et al.*, 2012; Geana *et al.*, 2016). Subsequently, these compositional parameters of the grape juice matrix can be used as predictors for the abovementioned factors and link to wine authenticity. Wine has become an important commodity world-wide and therefore, ensuring that imported wines are of a particular quality, and have not been illegally adulterated, is in the interest of producers, consumers and the relevant authorities (Geana *et al.*, 2016). However, the profiling of wines can be further complicated by the fermentation process (through the use of various conditions and strains of yeast and bacteria), aging, and storage conditions (Styger *et al.*, 2011).

The prediction of a grape variety based on compositional parameters of the grape must is less common. This is most likely due to the minimal economical relevance. However, the accurate

prediction of a grape must variety and origin based on a component of the grape juice matrix implies that the component is *characteristic* of that particular variety or origin. This information may aid the understanding of winemakers and viticulturists, and, subsequently, help them to make more informed decisions regarding practices and processes that could be employed to ensure the desired quality and style of the final wine.

Due to the central role of nitrogenous compounds in yeast metabolism and, consequently, the modulation of the organoleptic qualities of the resulting wine, knowledge of the amino acid profile, and how characteristic this profile is of a certain variety would be advantageous. This is especially relevant in terms of the direction that nitrogen research is currently moving in, whereby the specific nitrogen demand of various strains of *Saccharomyces cerevisiae* (Vilanova *et al.*, 2007) and non-*Saccharomyces* yeast (Rollero *et al.*, 2018) are being investigated. Therefore, this study aimed to elucidate the amino acid profile of an array of grapevine cultivars relevant to the South African wine industry and to investigate how well these cultivars could be predicted based on their amino acid profile, regardless of origin or vintage.

5.2. Materials and Methods

5.2.1. Sample Collection

The amino acid profile of 738 commercial grape juices was obtained over the 2016 and 2017 harvests. Samples were collected from various grape-growing districts across the Western Cape region of South Africa. All of the white cultivars' samples were collected as settled juices and red cultivars directly after crushing. All cultivars were harvested at a ripeness level suitable for commercial winemaking, according to the cellars participating in the survey. Samples were coded upon collection and stored at -20°C until analysis.

The survey followed an unsupervised format, resulting in the collection of 13 different cultivars, seven white and six red. The cultivars collected included: Cabernet Franc (n=13), Cabernet Sauvignon (n=38) Chardonnay (n=97), Chenin Blanc (n=176), Cinsaut (n=15), Grenache Blanc (n=17), Merlot (n=29), Pinotage (n=12), Roussanne (n=15), Sauvignon Blanc (n=219), Sémillon (n=16), Shiraz (n=51), and Viognier (n=40).

5.2.2. Amino Acid Analysis

Amino acids analysed were: alanine (Ala), arginine (Arg), aspartic acid (Asp), γ-amino butyric acid (GABA), glutamine (Gln), glutamic acid (Glu), glycine (Gly), histidine (His), hydroxyproline (Hyp), isoleucine (Ile), leucine (Leu), lysine (Lys), methionine (Met), ornithine (Orn), phenylalanine (Phe), proline (Pro), serine (Ser), threonine (Thr), tryptophan (Trp), and valine (Val).

Determination of individual amino acids was done using the AccQ-Tag Ultra amino acid kit (Waters), consisting of eluents A and B, the AccQ-Tag Ultra C_{18} column (2.1 x 100 mm, 1.7 µm) and a derivatization kit which comprised of AccQ-Tag derivatizing agent (6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC)), dry acetonitrile for preparing the AQC, and sodium borate buffer to be used in the derivatization reaction. The standard solution used was purchased from Thermo Scientific and it contained 2.5 µM/mL of each amino acid with the exception of cysteine which was at 1.25 µM/mL. Additional amino acids which were included in the analysis (tryptophan, ornithine, glutamine, and γ-amino butyric acid) were prepared initially as a stock solution of 2.5 µM/mL each. Norvaline (Nrv) was used as Internal Standard (IS). Stock solutions of standards were diluted from 1/2 to 1/1000 for the 11-point calibration (1250 nM/mL to 1.25 nM/mL). Sample preparation for both calibration standards and samples consisted of 800 µL sample to which 200 µL IS (200 ppm Nrv) were added and vortexed. Ten µL of this mixture, 70 µL of buffer and 20 µL of derivatization reagent were thoroughly mixed, followed by incubation at 55°C for 10 min and then placed in the autosampler tray of the instrument.

The instrumental analysis was performed on a Waters Acquity UPLC system with photodiode array (PDA) detector at 254 nm. Injection volume was 1 µL, analysis flow rate 0.7mL/min, and column temperature 60°C.

5.2.3. Statistical Analysis

Box plots were constructed using the statistical software SPPS (IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp). General discriminant analysis was performed using the statistical software package STATISTICA (version 13, TIBCO Software Inc. 2017, http://statistica.io).

5.3. Results and Discussion

This study had two major aims: (i) to provide insight into the amino acid composition of a range of grape varieties as well as (ii) to investigate the predictive ability of the amino acid profile in discriminating between the various cultivars included in the study. Therefore, the discussion will proceed by first describing the amino acid profiles by identifying the most abundant, as well as the least abundant amino acids for each variety. Subsequently, the potential significance of the respective amino acids is discussed in the context of the grapevine and fermentation. To address the second aim, the amino acid profile was first investigated for its ability to predict whether a particular variety was red or white. In addition to this, separate models for red and white cultivars were built to investigate whether the complete amino acid profile (from which subsets were selected) or proline, arginine and the proline/arginine ratio were better at identifying the specific variety.

5.3.1. Proline and arginine

Proline and arginine were found to be the most abundant amino acids, with an average of 697.69 mg/L for proline (range 33.22-3445.43 mg /L) and 388.35 mg/L for arginine (range 13.56-1616.56 mg/L) across all vintages, regions, and cultivars (Tables 5.1, B5.5, and B5.6). This is in agreement with previous studies surveying the amino acid content of *Vitis vinifera* varieties, where these amino acids were in most cases found to be orders higher than the rest (Kliewer, 1970; Huang & Ough, 1991; Spayd & Andersen-Bagge, 1996; Stines *et al.*, 2000). The large variation obtained is most likely due to the inclusion of different varieties and geographical origins across both vintages surveyed.

Early studies of amino acid profiles of grapevine have suggested that the proline to arginine ratio can be used as an index to discriminate between cultivars (Huang & Ough, 1991). Thus, future studies started profiling cultivars according to whether they were proline or arginine accumulators, with proline accumulators indicated by a ratio of >1 and arginine accumulators indicated by a ratio of <1. Merlot, Cabernet Sauvignon, and Chardonnay were found to be the highest proline accumulators, whereas Cinsaut, Pinotage, and Grenache Blanc were found to be the lowest proline accumulators (Figure 5.1). Sauvignon Blanc on the other hand was observed to have on average, equal concentrations of proline and arginine. These results are all similar to what has been found previously (Kliewer, 1970; Huang & Ough, 1991; Spayd & Andersen-Bagge, 1996; Stines *et al.*, 2000; Hannam *et al.*, 2016). However, although Merlot was found to have a high proline:arginine ratio, Huang and Ough (1991) and Spayd and Andersen-Bagge (1996) found Cabernet Sauvignon to have higher ratios than Merlot, contrary to the findings of the present study where Merlot had the highest proline to arginine ratio (proline:arginine 8.83). Interestingly, the study conducted by Huang and Ough (1991), also found a ratio of exactly one for Sauvignon Blanc in the 1987 vintage sampled. However, in the following year, Sauvignon Blanc grapes had higher arginine concentrations, possibly due to the different origin (Huang & Ough, 1991). Furthermore, Chardonnay was consistently found to be the white cultivar with the highest proline to arginine ratio (Kliewer, 1970; Huang & Ough, 1991; Spayd & Andersen-Bagge, 1996; Stines *et al.*, 2000; Hannam *et al.*, 2016).

Table 5.1. Overall mean, standard deviation, minimum mean, and maximum mean amino acid concentrations (mg/L).

Amino acid	Mean	Standard deviation	Min	Max
PRO	697.69	410.30	186.18	1638.23
ARG	388.35	168.26	185.63	765.56
GLN	111.57	41.83	61.06	216.37
TRP	105.67	55.27	47.33	214.29
GABA	100.18	19.92	69.21	136.60
ALA	85.17	34.91	38.55	145.08
SER	75.16	16.61	49.14	104.56
THR	70.27	24.30	35.49	116.10
GLU	61.12	18.65	35.23	100.96
HIS	31.07	8.64	21.11	51.92
VAL	28.81	5.43	20.52	37.74
PHE	27.62	12.19	10.09	49.75
ASP	25.69	8.44	11.98	37.86
ILE	19.92	5.89	13.93	32.78
LEU	16.07	8.03	5.38	32.02
HYP	10.85	3.48	2.79	16.91
LYS	3.91	1.27	2.32	6.07
MET	3.64	3.01	0.66	9.96
GLY	3.28	1.19	1.08	5.25
ORN	2.01	1.41	0.43	4.75

Bell and Henschke (2005), proposed that this ratio could also be used as an indicator of the ratio of assimilable nitrogen to non-assimilable nitrogen. This rule seems to hold up for most cultivars, for example, Grenache Blanc, Pinotage and Cinsaut are all high-YAN yielding cultivars (Chapter 3 Section 3.3.1), with a proline to arginine ratio of <1 and vice versa for cultivars such as Merlot, Cabernet Sauvignon and Cabernet Franc. On the other hand, this rule does not appear to apply for Chardonnay, a cultivar that is typically found to have very high average YAN concentrations (Butzke, 1998; Nicolini *et al.*, 2004; Hagen *et al.*, 2008). The ratio of proline to arginine as a cultivar indicator was, however, not found to be feasible by Spayd and Andersen-Bagge (1996) due to the large variation found in the juices surveyed.

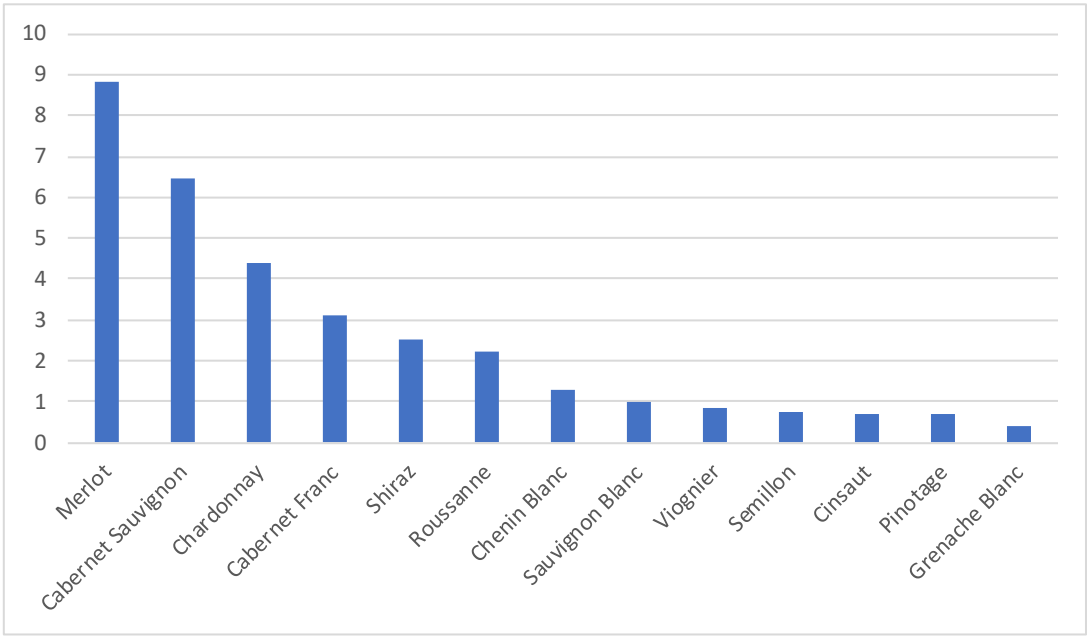


Figure 5.1. Proline/arginine ratio per cultivar, arranged in descending order.

Proportionally, for the amino acids quantified in this study, proline was observed to make up from 14.4% (Grenache Blanc) to 69.21% (Merlot) and on average, 35.7% of the grape juice amino acid content (Table B5.1). Moreover, proline contributed to approximately half (49.6%) and two-thirds (61.8%) of the amino content present in Cabernet Franc and Cabernet Sauvignon juices, respectively (Table B5.1). The similar proline content between Merlot, Cabernet Sauvignon and Cabernet Franc is not surprising due to the close genetic relationships exhibited between these cultivars (Myles *et al.*, 2011). On the other hand, arginine content ranged between 7.8% for Merlot to 34.82% for Grenache Blanc (Table B5.1). When arranging the percentage of arginine content from least to most and proline from most to least, in most cases the cultivars line-up or are relatively close to lining up (Figure 5.2). Therefore, it appears that proline and arginine concentrations are, to a degree, inversely proportional to one another.

Proline and arginine metabolism in the grapevine is linked via a common intermediate, ornithine. During arginine synthesis, ornithine is formed as an intermediate from glutamate, and again when arginine is broken down through arginase. Ornithine can, however, also result in the formation of P5C (Δ^1-pyrroline-5-carboxylate) which is a precursor of proline (Majumdar *et al.*, 2015). Thus, due to their linked metabolism, this inverse relationship between proline and arginine makes sense.

Proline accumulation in the grape berry occurs towards the end of ripening, typically in the last 4-6 weeks before harvest (Stines *et al.*, 1999). The reason for proline accumulation in plants is, however, a topic of debate. Most commonly, proline accumulation has been identified as a stress-response mechanism, protecting tissues against oxidative and osmotic stress. Another theory proposes that,

rather than being an adaptive mechanism, the accumulation of proline is a *product* of the stress imposed on the plant (Reviewed by: Ashraf & Foolad, 2007).

% Arginine		% Proline	
7,84%	Merlot	Merlot	69,21%
9,57%	Cabernet Sauvignon	Cabernet Sauvignon	61,79%
10,58%	Chardonnay	Cabernet Franc	49,59%
15,88%	Cabernet Franc	Chardonnay	46,36%
16,84%	Roussanne	Shiraz	44,56%
17,61%	Shiraz	Roussanne	37,10%
21,50%	Chenin Blanc	Chenin blanc	27,34%
24,55%	Sauvignon Blanc	Viognier	25,95%
29,51%	Cinsaut	Sauvignon Blanc	24,56%
30,34%	Sémillon	Sémillon	22,08%
30,42%	Pinotage	Pinotage	20,79%
30,65%	Viognier	Cinsaut	20,71%
34,82%	Grenache Blanc	Grenache Blanc	14,42%

Figure 5.2. Percentage of amino acids per cultivar contributed by arginine, arranged in ascending order and percentage of amino acids contributed by proline, arranged in descending order.

However, Stines *et al.* (1999) found that the accumulation of proline in the grape berry is independent from the stress-induced pathway and that the accumulation is essentially a part of normal fruit development. This was hypothesised due to the findings that the proline accumulation occurring in developing berries were not regulated by fluctuations in P5CS (Δ^1-pyrroline-5-carboxylate synthetase) mRNA or protein levels associated with proline biosynthesis from glutamate, or by changes in the levels of PDH (proline dehydrogenase) – proteins which are related to the breakdown of proline. Therefore, other regulatory mechanisms were thought to be involved. However, the synthesis of proline from ornithine through the OAT (ornithine δ-aminotranferase) pathway could also not be confirmed in this study. Thus, Stines *et al.* (1999), does not provide conclusive evidence as to the mechanisms involved in proline accumulation during berry development, although they refute the stress-related hypothesis. Furthermore, in a later study by Stines and colleagues, investigating the accumulation of proline and arginine in grape berries in relation to berry maturity, tissue type and cultivar, it is argued that stress-induced proline accumulation does not at all occur in grape berries (Stines *et al.*, 2000). They support this hypothesis by studies done on partial root drying and deficit irrigation techniques employed to enhance water usage efficiency by the grapevine. They

report that the levels of water stress investigated in these studies were not observed to significantly impact the levels of free proline in the berries (Mccarthy, 1997; Loveys *et al.*, 2000). However, a literature search into this topic has shown that there are indeed studies that found increased proline accumulation due to osmotic stress, induced by water-deficit irrigation techniques (Cramer *et al.*, 2013; Romero *et al.*, 2015).

Therefore, the reasons for proline accumulation remains a controversy. The contradictory findings of these various studies – as well as the cultivar specific levels of proline that are observed in both this survey as well as previous investigations, may indicate that proline accumulation cannot be generalised across all cultivars. Furthermore, proline accumulation may be a more intricate interaction between the genetics of the vine and the particular set of environmental conditions, than what may have previously been thought.

5.3.2. Abundant amino acids

Other than proline and arginine, on average, glutamine (111.57 mg/L), tryptophan (105.67 mg/L), γ-amino butyric acid (GABA, 100.18 mg/L), and alanine (85.17 mg/L) were found to be the most abundant amino acids (Table 5.1). Furthermore, not including proline or arginine, these amino acids were found to be the four most abundant amino acids for each of the cultivars surveyed, appearing in varying orders of abundance. Glutamine was found to be the third most abundant amino acid in Cinsaut, Grenache Blanc, Pinotage, Sauvignon Blanc, Semillon and Viognier. In Cabernet Franc and Roussanne, tryptophan was found to be the third most prevalent, while for Cabernet Sauvignon and Merlot, it was GABA. Finally, alanine was found to be the third most predominant amino acid for Chardonnay and Chenin Blanc juices (Figures 5.3 and 5.4).

Huang and Ough (1991) also reported glutamine to be one of the most prevalent amino acids. Glutamine is central to amino acid metabolism in the grapevine, as it is the primary form of transportable nitrogen through the phloem and into the berry, and secondly, is the precursor molecule for an array of other amino acids. Therefore, during the early stages of berry development, this amino acid is found to be the most abundant. However, during the later stages of development, there is a marked decline in concentration which is due to the conversion into other amino acids such as glutamate, proline, ornithine, and arginine (Stines *et al.*, 2000). In the current study, glutamine concentrations were observed to make up on average 6.1% of the amino acid content for the cultivars surveyed (Table B5.1). Furthermore, on average, Merlot was observed to have the lowest glutamine concentrations (61.06 mg/L), and Pinotage the highest (216.37 mg/L) (Table 5.1). Proportionally, Grenache Blanc was found to have the highest glutamine content with an average of 9.9% of the total amino acid concentration being contributing by this amino acid (Table B5.1).

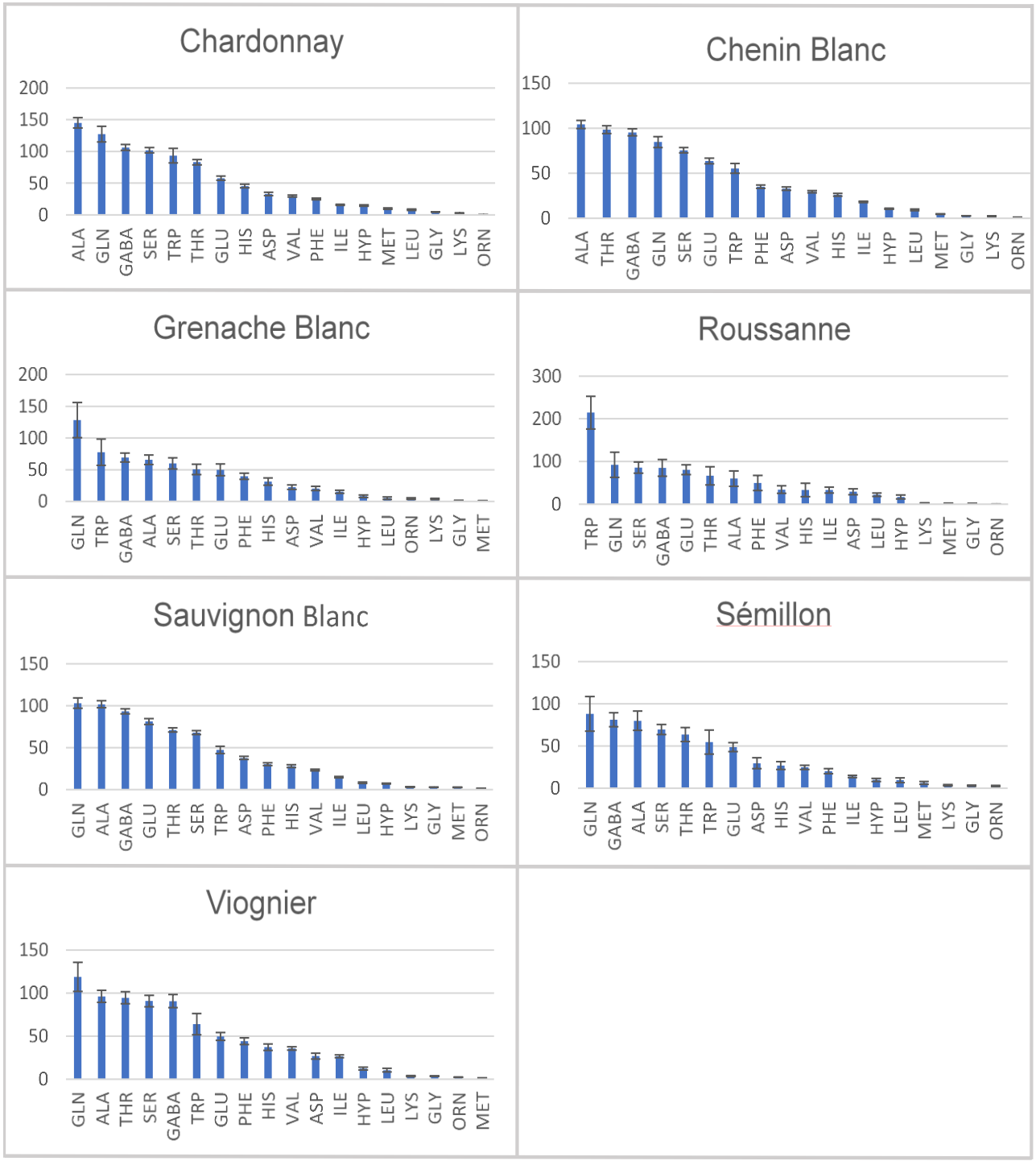


Figure 5.3. Mean concentrations of amino acids, excluding proline and arginine, of white cultivars included in the survey, arranged from most to least. Error bars indicate the standard error.

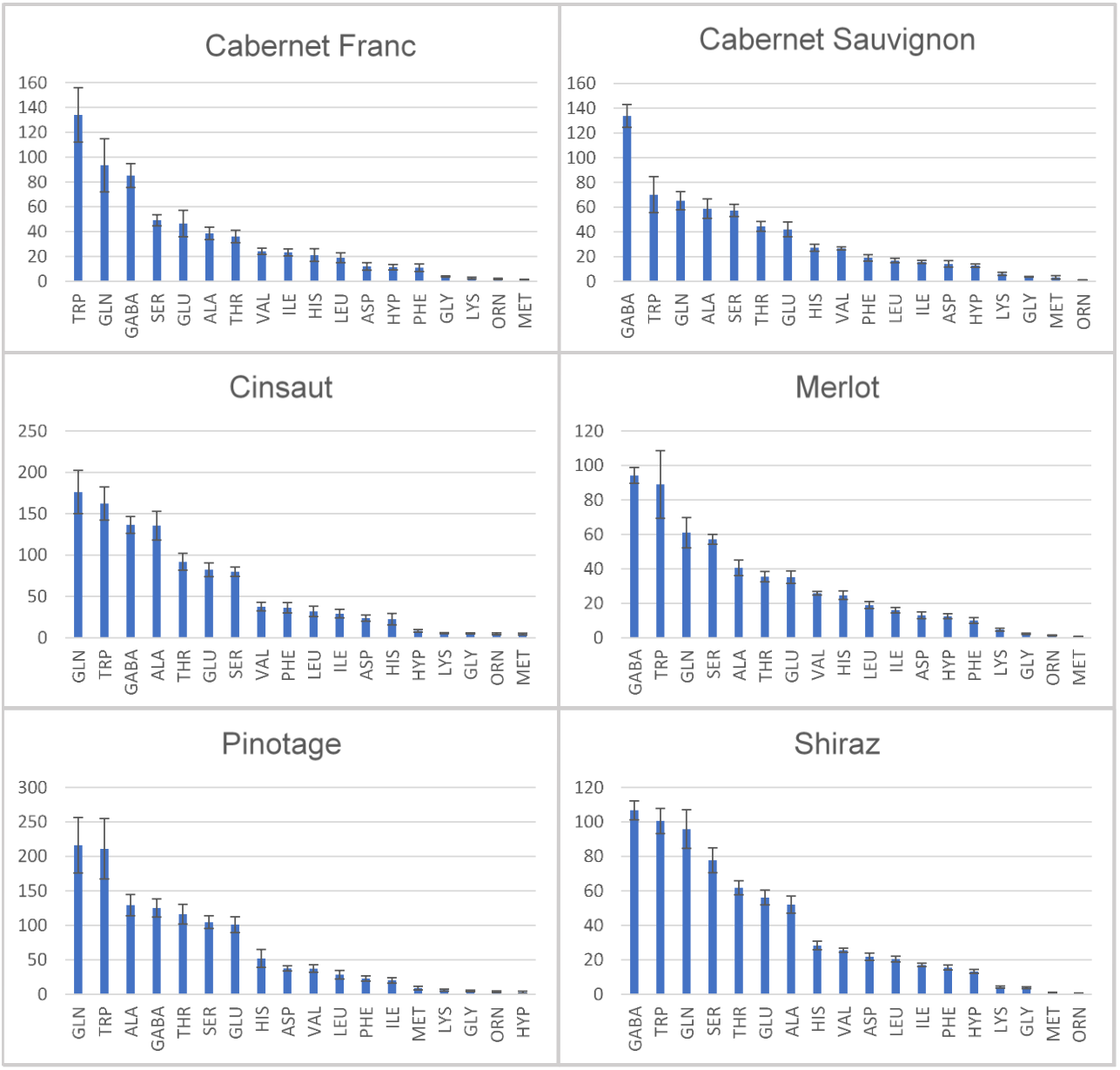


Figure 5.4. Mean concentrations of amino acids, excluding proline and arginine, of red cultivars included in the survey, arranged from most to least. Error bars indicate the standard error.

Alanine has also previously been identified as an amino acid occurring in high concentrations, regardless of the cultivar, origin, or vintage (Kliewer, 1970; Huang & Ough, 1991; Spayd & Andersen-Bagge, 1996; Stines *et al.*, 1999). Chenin Blanc and Sauvignon Blanc juices from Washington from 1986 to 1990 had alanine as the third most predominant amino acid after proline and arginine (Spayd & Andersen-Bagge, 1996). Although, in the current study, glutamine was found to be, on average, the third most abundant amino acid in Sauvignon Blanc (103.19 mg/L), alanine still followed closely with an average of 101.88 mg/L (Table B5.2). Furthermore, the current study found Chardonnay to be the highest alanine-containing cultivar, having an average concentration of 145.08 mg/L (Table B5.2, Figure 5.3). Percentage-wise, alanine was found to make up approximately 7% of the total

amino acid content of the aforementioned cultivars (Chardonnay: 6.9%; Chenin Blanc: 7.1%; and Sauvignon Blanc: 7.1%) (Table B5.1).

The quantification of GABA was reported by Kliewer (1970), Stines *et al*. (2000) and Asensio, Valdés, and Cabello (2002). In the current survey, average GABA concentrations were found to range between 69.21 mg/L for Grenache Blanc and 136.60 mg/L for Cinsaut (Table 5.1). The overall average was found to be 100.18 mg/L, making up approximately 5% of the grape juice amino acid content (Table 5.1 and Table B5.1). Cabernet Sauvignon, Chardonnay, Cinsaut, Pinotage and Shiraz were all found to have average GABA concentrations of more than 100 mg/L (Figure 5.3 and 5.4; Table B5.2). GABA is a non-proteinogenic amino acid and therefore, does not play a role in the formation of biomass, but rather in the regulation of plant growth and adaption to various forms of biotic and abiotic stress. Interestingly, a study conducted by Saloua et al. (2014), found elevated levels of GABA in combination with the upregulation of genes associated with the enzyme activities of polyamine oxidases in Meski, a drought resistant species of *Vitis vinifera*. This correlation points to the link between polyamine homeostasis and GABA formation and the subsequent increased tolerance of the vine towards drought conditions. Specifically, GABA production can occur due to the catabolism of polyamines through the enzymatic action of diamine oxidase (Agudelo-Romero *et al.*, 2013). It is through this catabolic process that grape berry concentrations of GABA (along with arginine) were observed to increase during ripening in a study on the metabolic profiling of the varieties Touriga Nacional, Aragones, and Trincadeira (Ali *et al.*, 2011).

Tryptophan, found as the third most abundant amino acid in Roussanne (214.29 mg/L) and Cabernet Franc (211.11 mg/L) (Figure 5.3 and 5.4; Table B5.2), is a member of the aromatic amino acid family. This amino acid is particularly important along with the other aromatic (phenylalanine and tyrosine) and branched-chain amino acids (isoleucine, leucine, and valine) in the formation of favourable fruity and floral aromas during fermentation. However, the importance of tryptophan in the grapevine stems from its role in the production of auxin which is a hormone which plays a pivotal role in berry ripening (Böttcher *et al.*, 2013).

Due to interference of the derivatization agent with tyrosine, this aromatic amino acid could not be accurately quantified and was thus not included in the calculations. Despite this, aromatic amino acids were observed to make up a larger proportion of the total amino content for each cultivar – on average 7.1% of the amino acid content – compared to 3.5% for the branched-chain amino acids (Table 5.2). Proportionally, Cabernet Sauvignon and Merlot had the lowest aromatic amino acid content, with only 4.1% and 4.2%, respectively (Table 5.2). These cultivars were also observed to have amongst the lowest proportions of branched-chain amino acids (Cabernet Sauvignon 2.8% and Merlot 2.6%), along with Chardonnay 2.6%. Moreover, Roussanne was found to have, on average, the highest aromatic amino acid content, in both absolute terms (264 mg/L) and proportionally (13.4%) (Table 5.2).

In terms of the branched-chain amino acids, Cinsaut and Roussanne were found to have the highest proportions - with 4.6% and 4.5%, respectively (Table 5.2). Therefore, as Roussanne contains high concentrations of these precursor molecules (both aromatic and branched-chain amino acids), it can be identified as a cultivar with a great amount of aromatic potential in terms of the production of fusel alcohols and esters. However, these positive aroma compounds are only produced when the total YAN concentration is capable of fulfilling the full biosynthetic requirement of the yeast. As Roussanne has been identified as a cultivar which has a very low total YAN content (average 132 ± 34 mg N/L; Chapter 3, Section 3.3.1), it will most likely require nutrient supplementation in the form of DAP or complex nutrients to realise its full aromatic potential.

Table 5.2. Average concentration and percentage of branched-chain and aromatic amino acids per cultivar.

	Concentration (mg/L)		Percentage	
Cultivar	**Branched-chain**	**Aromatic**	**Branched-chain**	**Aromatic**
Cabernet Franc	66.4	144,9	3.7	8.2
Cabernet Sauvignon	59.2	89,1	2.7	4.1
Chardonnay	53.8	118,5	2.6	5.6
Chenin Blanc	57.2	90,3	3.9	6.2
Cinsaut	99.0	198,6	4.6	9.2
Grenache Blanc	41.3	117,3	3.2	9.1
Merlot	60.9	99,1	2.6	4.2
Pinotage	85.7	234,0	3.4	9.3
Roussanne	88.8	264,0	4.5	13.4
Sauvignon Blanc	46.3	77,8	3.2	5.5
Semillon	48.2	75,0	3.6	5.6
Shiraz	63.0	116,1	3.4	6.3
Viognier	72.5	108,0	3.9	5.8
Overall average	133.3	64,8	3.5	7.1

5.3.3. Least abundant amino acids

Ornithine (2.01 mg/L), glycine (3.28 mg/L), methionine (3.64 mg/L) and lysine (3.91 mg/L) were found to have the lowest concentrations, both in terms of the overall average, as well as per cultivar (Table B5.2). This is again in agreement with what has been published previously (Huang & Ough, 1991; Spayd & Andersen-Bagge, 1996; Stines *et al.*, 2000). The low concentration of ornithine is most likely due to its central role in nitrogen metabolism, acting as a precursor molecule for the formation of the most abundant amino acids, arginine and proline, as well as its involvement in polyamine synthesis through ornithine decarboxylase. As *Saccharomyces cerevisiae,* the principal yeast used for fermentation, is not able to efficiently metabolise glycine and lysine, these amino acids are considered as a poor source of nitrogen for this yeast (Garde-Cerdán *et al.*, 2009; Jolly *et al.*, 2017).

However, these amino acids may be assimilated by some non-*Saccharomyces* yeasts (Rollero *et al.*, 2018). There is currently increasing research into the benefits of allowing the growth and fermentation activity of non-*Saccharomyces* yeasts on the organoleptic quality and complexity of the final wine (Comitini *et al.*, 2017; Gobert *et al.*, 2017; Jolly *et al.*, 2017; Rollero *et al.*, 2018).

Cysteine, a sulfur-containing amino acid present in grape juice, could not be accurately quantified and is a shortcoming of the analytical method employed in this study. Aside from the very low concentration of this compound in the grape juice matrix (Spayd & Andersen-Bagge, 1996), this is thought to be due to the reactivity of the S-H group together with the interference of the high sugar matrix. Cysteine is particularly important in the context of winemaking as a deficiency in this amino acid, along with other S-containing amino acids may lead to off-flavour production (H_2S). Furthermore, the varietal aromas of cultivars such as Sauvignon Blanc, Sémillon and Riesling are due to volatile thiols ((4-mercapto-4-methylpentan-2-one (4MMP); 4-mercapto-4-methylpentan-2-ol (4MMPOH) and 3-mercaptohexan-1-ol (3MH)) which occur as odourless cysteine-conjugates and are converted into their volatile state by action of the yeast during fermentation. Therefore, it is not surprising that Spayd and Andersen-Bagge (1996) found cysteine concentrations to be above the detection threshold in the cultivars Sauvignon Blanc and Riesling.

5.3.4. Overall view of the amino acid profiles

A heatmap of the relative average amino acid concentrations presents the z-score indicating how much (in terms of standard deviations) the average amino acid concentration per cultivar deviates from the overall average across all cultivars (Figure 5.5). This method of representing the data provides a comprehensive overview of how the cultivars may compare to one another based on their amino acid content. Furthermore, the associated dendrogram indicates how the cultivars may *relate* to one another based on their amino acid profiles. Therefore, when looking at the heatmap horizontally, the amino acid profile per cultivar can be observed, whereas vertically, the relative average concentrations can be compared across cultivars for a specific amino acid. Thus, in this representation, the cultivars containing very high or very low concentrations (in comparison to the mean) can be identified. For example, the white cultivars, Grenache Blanc, Sémillon, Sauvignon Blanc, and Chenin Blanc appear to group together based on the lower concentrations of amino acids compared to the other cultivars included in this study. Furthermore, it is clear that Merlot is the cultivar with the highest concentration of proline and that Pinotage – and to a lesser degree, Cinsaut – generally has higher concentrations of most of the amino acids compared to the other cultivars surveyed. The close genetic relationship between Pinotage and Cinsaut together with the similarity in the amino acid profile highlights the influence of the genetic make-up in determining the grape must composition.

With good reason, in the past three decades, a great deal of emphasis has been placed on ensuring that the *total* YAN concentration is adequate to support sufficient biomass production, thereby

avoiding stuck fermentations (Bely *et al.*, 1990b; Henschke & Jiranek, 1993; Bisson, 1999). However, more recently, the relevance of the *content* of the nitrogen sources in *relation* to one another in determining the overall style and quality of the final wine is emerging. This is due to the complex metabolic activities of the yeast where the ratio of amino acids to another will determine the flux of these nitrogenous compounds into the various metabolic pathways of the yeast, and subsequently, influence the organoleptic qualities produced (Beltran *et al.*, 2004; Gobert *et al.*, 2017; Rollero *et al.*, 2018). Thus, when a grape must is supplemented with nitrogenous compounds, it is not only about increasing the nitrogen content *per se*, but about how this increase alters the *ratio* of amino acids (and ammonia) to another.

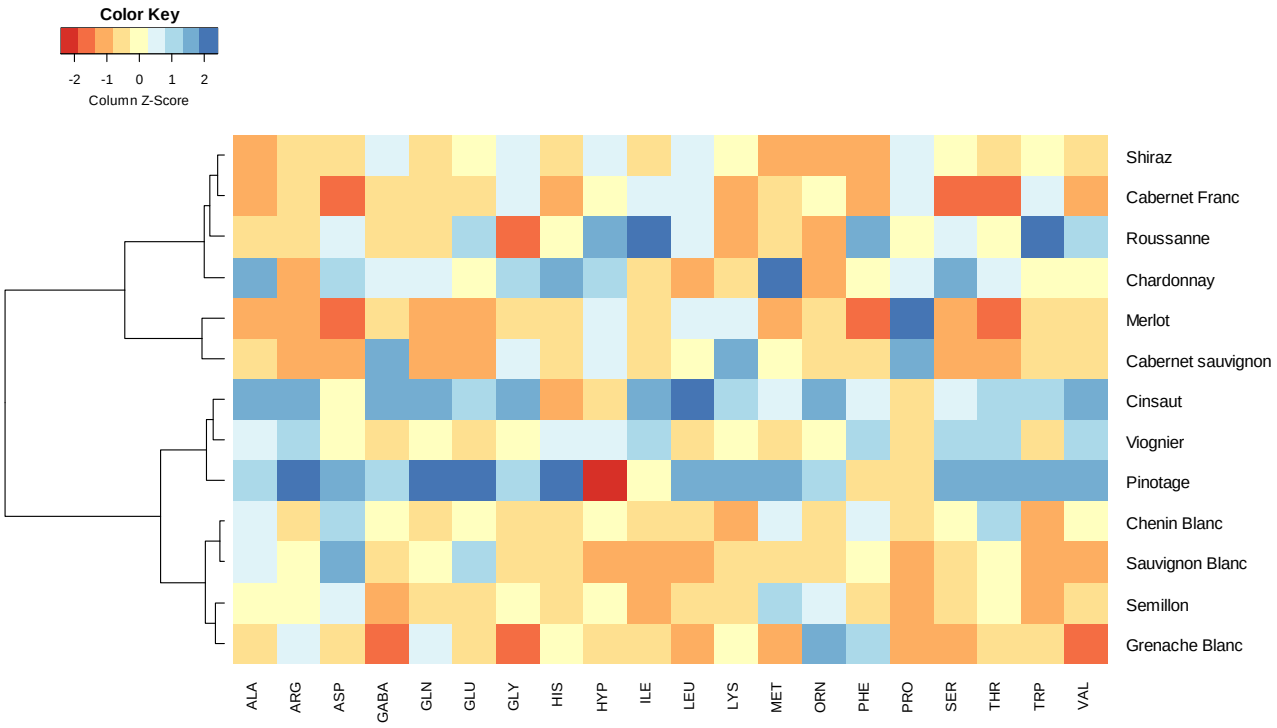


Figure 5.5. Heatmap of the average amino acid concentrations and dendrogram illustrating how these cultivars relate to one another based on these average concentrations.

5.3.5. Predictive ability of the grape must amino acid profile

The data was also used to evaluate how accurately the amino acid composition could be used to discriminate between cultivars and predict a certain cultivar. This was achieved using General Discriminant Analysis (GDA), a modelling technique involving the application of the general linear model (GLM) algorithm to the discriminant analysis function. The benefits of this include the possibility of a “best-subset” selection criteria. The optimum number of predictors are selected based on leave-one-out cross-validation. The best subset is then subsequently selected based on how many times the predictor appears in the 20 best models. Whether or not the predictor variable is statistically significant was tested by the Wilk’s Lambda statistic (Tables B5.9, B5.10, and B5.11).

5.3.5.1. Discrimination between red and white varieties

As a first step, for the discrimination between white and red grape juices, alanine (19), leucine (18), GABA (17), and proline (6) were the amino acids that achieved the best prediction (Table 5.3). Numbers included in brackets are the number of times the amino acid appeared in the best 20 models. The training set included 517 samples (111 red and 406 white) *i.e.* 70% of the data. Thus, the models were independently validated with the remaining 30% of the data. Overall, the model predicted 82.8% of samples correctly. When looking at the misclassification table, only 66% of the red grape juice samples were correctly predicted whereas 87% of white samples were correctly predicted (Table 5.4). However, the decline in the performance of the model in distinguishing between red and white samples is hypothesised to be due to the markedly lower number of red samples included in the study.

Due to the lower number of samples contributed by each of the individual red cultivars compared to the white cultivars, class membership of *specific* cultivars was predicted in two separate models, one for white and one for red.

Table 5.3. The best subset of predictor variables that were identified through general discriminant analysis for the prediction of white and red cultivars and the number of times these predictor variables occurred in the 20 best models.

Best Subset	Number of times AA appears
Red vs. White	
ALA	19
LEU	18
GABA	17
PRO	6
White	
ARG	20
MET	20
THR	18
PRO	17
ALA	15
GLU	9
Red	
GABA	20
PRO	19
PHE	17
HYP	12
THR	11
ILE	6

Table 5.4. Misclassification table and overall percentage of white and red cultivars correctly predicted based on the best subset principal.

Best subset	Percent Correct	Red	White
Red	66	31	16
White	87	22	152
Total	**75.6**	**53**	**168**

5.3.5.2. Prediction of white cultivars

White cultivars: Best subset

Due to the vast difference in the number of samples of white cultivars, only cultivars with more than 30 samples were included in this discriminatory analysis. Thus, Chardonnay (n=97), Chenin Blanc (n=179), Sauvignon Blanc (n=219) and Viognier (n=40) were considered. Training to test set ratios were again randomly divided into a 70/30 ratio. This meant that of the 532 samples included in the analysis, a test set of 160 samples was used to independently validate the model. Alanine (15) and proline (17) were again included in the best subset, in addition to arginine (20), methionine (20), threonine (18), and glutamic acid (9) (Table 5.3). This model was able to correctly identify 75.6% of the white grape juice samples according to cultivar (Table 5.5). Furthermore, these results confirm the results observed in the heatmap, where Sauvignon Blanc, Chenin Blanc, and Viognier were found to be more similar to one another than any of these cultivars were to Chardonnay (Figure 5.5). Specifically, a 100% of Chardonnay, 73.6% of Chenin Blanc, 65.2% of Sauvignon Blanc and 83.3% of Viognier samples were correctly predicted (Table 5.5). Sauvignon Blanc had the lowest prediction accuracy and was mainly misclassified as Viognier (15%) and Chenin Blanc (14%).

Table 5.5. Misclassification table and overall percentage of white cultivars correctly predicted based on the best subset principal

Best subset	Percent Correct	Chardonnay	Chenin Blanc	Sauvignon Blanc	Viognier
Chardonnay	100.0	29	0	0	0
Chenin Blanc	73.6	2	39	7	5
Sauvignon Blanc	65.2	4	9	43	10
Viognier	83.3	0	1	1	10
Total	**75.6**	**35**	**49**	**51**	**25**

White cultivars: Proline and arginine

The use of proline and arginine as the only predictor variables had a markedly lower predictive ability even though these predictors were both found to be statistically significant according to the Wilk's lambda statistic ($p < 0.05$) (Table B5.10). The overall predictive ability of the model decreased from 76% (using the best subset) to 47.5% by using only the average concentration of proline and arginine (Table 5.6). The same trends were observed for this model as with the models based on the best subset where Chardonnay was found to be the most accurately predicted cultivar (75.9%) and Sauvignon Blanc the most poorly predicted cultivar (21.2%).

When adding the ratio of proline to arginine (proline/arginine) as a predictor variable to this model, the overall performance did not change (47.5%) (Table 5.7), however, the predictive ability of specific cultivars did vary. Adding this ratio increased the predictive ability of Chardonnay (82.8%), however, it led to a decrease in the prediction accuracy of Sauvignon Blanc (16.7%). The prediction accuracy of Chenin Blanc was marginally improved (66%), whereas Viognier remained unchanged at 50%. Sauvignon Blanc was in both instances (where just the average concentrations of proline and arginine were used as predictor variables as well as in the case of the addition of the proline/arginine ratio), most often misclassified as Chenin Blanc. The misclassification of Sauvignon Blanc as Chenin Blanc may stem from the close genetic relationship exhibited between these cultivars (Myles *et al.*, 2011). Furthermore, the poor prediction of Sauvignon Blanc may also be due to the large number of diverse sample types that were collected (Chapter 3, Section 3.3.1).

Therefore, using proline and arginine as the sole predictor variables may offer a small degree of differentiation between certain cultivars such as Chardonnay from Chenin Blanc, Sauvignon Blanc and Viognier but cannot definitively be used as an indicator to discriminate between cultivars as hypothesised by Huang and Ough (1991).

Table 5.6. Misclassification table and overall percentage of white cultivars correctly predicted based on the average proline and arginine concentrations as predictor variables.

Proline + Arginine	Percent Correct	Chardonnay	Chenin Blanc	Sauvignon Blanc	Viognier
Chardonnay	75.9	22	7	0	0
Chenin Blanc	64.2	2	34	6	11
Sauvignon Blanc	21.2	4	32	14	16
Viognier	50.0	0	0	6	6
Total	**47.5**	**28**	**73**	**26**	**33**

Table 5.7. Misclassification table and overall percentage of white cultivars correctly predicted based on the average proline and arginine concentrations as well as the ratio of proline/arginine as predictor variables.

Proline + Arginine + Proline/Arginine	Percent Correct	Chardonnay	Chenin Blanc	Sauvignon Blanc	Viognier
Chardonnay	82.8	24	5	0	0
Chenin Blanc	66.0	4	35	8	6
Sauvignon Blanc	16.7	2	36	11	17
Viognier	50.0	0	0	6	6
Total	**47.5**	**30**	**76**	**25**	**29**

5.3.5.3. Prediction of red cultivars

Red cultivars: Best subset

As less samples of red cultivars were collected during the survey, all red cultivars included in the amino acid survey were included in the model. Therefore, the cultivars considered included Cabernet Franc (13), Cabernet Sauvignon (38), Cinsaut (15), Merlot (29), Pinotage (12), and Shiraz (51). Overall, the model correctly predicted 60.1% of the red grape juice samples according to cultivar (Table 5.8). However, due to the lower number of samples, cross-validation was used instead of an independent test-set to validate the model. Pinotage was most frequently correctly identified (75%), with only 1 sample being misclassified as Shiraz and 2 as Cinsaut. The misclassification of Pinotage as Cinsaut may also stem from their close genetic (parent-offspring) relationship. Furthermore, Cabernet Franc was most frequently misclassified as Merlot, also possibly due to the close genetic (parent-offspring) relationship exhibited between these cultivars. Moreover, even though Shiraz contributed the greatest number of samples to this data set, only 54.9% of the samples were correctly predicted. In addition to this, other cultivars were most often misclassified as Shiraz. Therefore, Shiraz appears to have an amino acid profile which is quite similar to the other cultivars included in the model and thus, not easily distinguishable (Table 5.8).

Table 5.8. Misclassification table and overall percentage of red cultivars correctly predicted based on the best subset principal.

Best Subset	Percent Correct	Cabernet Franc	Cabernet sauvignon	Cinsaut	Merlot	Pinotage	Shiraz
Cabernet Franc	61.5	8	0	0	4	0	1
Cabernet Sauvignon	60.5	2	23	3	4	1	5
Cinsaut	40.0	1	0	6	0	1	7

Table 5.8. (cont.)

Best Subset	Percent Correct	Cabernet Franc	Cabernet sauvignon	Cinsaut	Merlot	Pinotage	Shiraz
Merlot	72.4	2	3	2	21	0	1
Pinotage	75.0	0	0	2	0	9	1
Shiraz	54.9	7	7	0	3	6	28
Total	**60.1**	**20**	**33**	**13**	**32**	**17**	**43**

Red cultivars: Proline and arginine

Using proline and arginine as the only predictor variables for red cultivars also led to a distinctly lower predictive ability, where the overall model was only able to correctly predict 32.3% of red grape juice samples according to cultivar (Table 5.9). Cabernet Franc was the most poorly predicted cultivar with only 7.7% of its samples correctly identified. Cabernet Sauvignon and Cinsaut were also very poorly predicted, with only 15.8% and 13.3% of samples correctly identified, respectively. On the other hand, 69% of Merlot samples were correctly predicted and was therefore found to be the cultivar most accurately predicted based on average proline and arginine concentrations (Table 5.9). This is not surprising as this cultivar was found to have the most extreme concentrations for both of these amino acids – having the lowest arginine and the highest proline concentrations. Therefore, these two amino acids may be reasonably accurate to distinguish Merlot from other cultivars but may not be a good indicator overall.

The addition of the ratio of proline to arginine increased the prediction accuracy of the model by approximately 10% to 42.4% (Table 5.10). Although this model was not as accurate as the same model to predict white cultivars, the addition of the ratio made a bigger impact on the overall prediction accuracy of red cultivars, whereas the overall prediction accuracy of white cultivars remained the unchanged at 47.5%. The increased performance of this model was, however, owed to the improved prediction of Shiraz samples which increased from 31.4% (using only the average proline and arginine concentrations) to 62.7% (with the addition of the proline/arginine ratio). Interestingly, this model allowed for better prediction of Shiraz samples than the model using the best subset principal (54.9%) (Table 5.8). Furthermore, the prediction of Cabernet Franc, Cinsaut and Pinotage remained unchanged with the addition of the proline/arginine ratio as a predictor variable. However, the prediction accuracy of Merlot was observed to drop by 3.5% from 69% to 65.5% (Table 5.10).

Therefore, using proline and arginine concentrations as the sole predictor variables, as well as the addition of the ratio of these amino acids also yielded unsatisfactory results to distinguish red cultivars from one another.

Table 5.9. Misclassification table and overall percentage of red cultivars correctly predicted based on the average proline and arginine concentrations as predictor variables.

Proline + Arginine	Percent Correct	Cabernet Franc	Cabernet sauvignon	Cinsaut	Merlot	Pinotage	Shiraz
Cabernet Franc	7.7	1	3	2	3	0	4
Cabernet Sauvignon	15.8	4	6	1	16	1	10
Cinsaut	13.3	1	0	2	0	6	6
Merlot	69.0	1	3	1	20	1	3
Pinotage	50.0	1	0	3	0	6	2
Shiraz	31.4	21	1	5	4	4	16
Total	**32.3**	**29**	**13**	**14**	**43**	**18**	**41**

Table 5.10. Misclassification table and overall percentage of red cultivars correctly predicted based on the average proline and arginine concentrations as well as the ratio of proline/arginine as predictor variables.

Proline + Arginine + Proline/Arginine	Percent Correct	Cabernet Franc	Cabernet sauvignon	Cinsaut	Merlot	Pinotage	Shiraz
Cabernet Franc	7.7	1	3	4	3	0	2
Cabernet Sauvignon	18.4	3	7	1	14	1	12
Cinsaut	13.3	0	0	2	0	6	7
Merlot	65.5	2	4	0	19	1	3
Pinotage	50.0	1	0	3	0	6	2
Shiraz	62.7	7	1	3	4	4	32
Total	**42.4**	**14**	**15**	**13**	**40**	**18**	**58**

5.4 Conclusion

The nitrogen content of the grape berry is essential for the proper development and functioning of the grapevine – whether it be an adaptive mechanism to stress (such as in the case of proline and GABA) or whether it be involved in central nitrogen metabolism and berry ripening (as in the case with glutamine and tryptophan, respectively). However, when the grapes are harvested, the nitrogen content – specifically the yeast assimilable nitrogen portion – becomes important in the context of yeast metabolism and subsequently, the fermentation process.

Therefore, the current study aimed to provide a strong foundation for nitrogen research in the context of the wine industry. This was done by providing not only the *absolute* values but also the *proportions* of various amino acids for a range of industrially relevant cultivars as this will help to provide a more comprehensive picture of the potential of the grape must in terms of both fermentation efficiency and aroma. After proline and arginine, glutamine, tryptophan, GABA and alanine were found to be the most abundant amino acids. Ornithine, glycine, methionine and lysine were found to have the lowest overall concentrations, both on average as well as per cultivar. Merlot and Cabernet Sauvignon were found to have the lowest proportions of aromatic and branched-chain amino acids, with Roussanne being found to have the highest proportion of these precursors of fruity and floral aromas.

Therefore, as Cabernet Sauvignon and Merlot have been found to have very low total YAN concentrations, these cultivars would in most cases require nitrogen supplementation to ensure the completion of fermentation. However, the addition of complex nutrients (which may contain varying concentrations of these branched-chain and aromatic amino acids) may be a more beneficial supplementation strategy for these cultivars compared to ammonia addition (in the form of diammonium phosphate). On the other hand, as Roussanne already has high concentrations of these precursor molecules, the addition of (cheaper) ammonium may be sufficient to ensure not only the completion of fermentation but to ensure the formation of favourable organoleptic qualities in the final wine.

In addition to this, it was investigated how *characteristic* the amino acid profile is of a particular group (red or white) or of a particular cultivar. This was done by examining how accurately cultivars could be predicted based on their average amino acid concentrations using general discriminant analysis (GDA) and the best subset principal. Based on this, Chardonnay showed the highest prediction accuracy with a 100% of its samples correctly identified with regards to the white cultivars and Pinotage (75%) with regards to the red cultivars. Overall, the white cultivars included in this study were more accurately distinguished from one another (75.6%) compared to the red (60.1%). This predictive ability was subsequently compared to the accuracy of predicting cultivars based on only their arginine and proline concentrations as well as the ratio between the two, based on the findings by Huang and Ough (1991) who eluded to the potential of these amino acids to distinguish between

cultivars. The use of only these amino acids as well as the addition of the proline/arginine ratio as a predictor variable did not offer satisfactory discriminatory power between either white or red cultivars.

However, still the discrimination between white cultivars was found to be more accurate for the models including the use of only proline and arginine, as well as the addition of proline/arginine as predictor a variable, than it was between red cultivars. This is hypothesised to be because of the closer genetic relationships between the group of red cultivars included in this study than between the white cultivars (Myles *et al.*, 2011).

Therefore, general discriminant analysis using the best subset principal was able to provide reasonable predictive power and thus, there is merit in using amino acid profiles to distinguish between cultivars. However, prediction accuracy seemed to depend on, to a certain degree, how related cultivars were to one another.

To our knowledge, this was the first study to use the amino acid profile of such a large number of grape juice samples to discriminate between various cultivars. The possibility of this has only been eluded to by previous authors, especially for the prediction of white cultivars. Furthermore, this study tested the hypothesis of Huang and Ough (1991) who theorised that the proline and arginine concentrations as well as the ratio (proline/arginine) can be used as an indicator of cultivar.

References

Agudelo-Romero, P., Bortolloti, C., Pais, M.S., Tiburcio, A.F., Fortes, A.M., 2013. Study of polyamines during grape ripening indicate an important role of polyamine catabolism. Plant Physiology and Biochemistry 67, 105–119.

Ali, K., Maltese, F., Fortes, A.M., Pais, M.S., Choi, Y.H., 2011. Monitoring biochemical changes during grape berry development in Portuguese cultivars by NMR spectroscopy. Food Chemistry 124(4), 1760–1769.

Asensio, M., Valdés, E., Cabello, F., 2002. Characterisation of some Spanish white grapevine cultivars by morphology and amino acid analysis. Scientia Horticulturae 93, 289–299.

Ashraf, M. & Foolad, M.R., 2007. Roles of glycine betaine and proline in improving plant abiotic stress resistance. Environmental and Experimental Botany 59(2), 206–216.

Bell, S.-J. & Henschke, P.A., 2005. Implications of nitrogen nutrition for grapes, fermentation and wine. Australian Journal of Grape and Wine Research 11(3), 242–295.

Beltran, G., Novo, M., Rozès, N., Mas, A., Guillamón, J.M., 2004. Nitrogen catabolite repression in *Saccharomyces cerevisiae* during wine fermentations. FEMS Yeast Res 4(6), 625–632.

Bely, M., Sablayrolles, J.M., Barre, P., 1990a. Description of Alcoholic Fermentation Kinetics: Its Variability and Significance. Am J Enol Vitic. 41(4), 319–324.

Bely, M., Sablayrolles, J.-M., Barre, P., 1990b. Automatic detection of assimilable nitrogen deficiencies during alcoholic fermentation in oenological conditions. Journal of Fermentation and Bioengineering 70(4), 246–252.

Bisson, L.F., 1999. Stuck and Sluggish Fermentations. Am J Enol Vitic. 50(1), 107-119.

Böttcher, C., Burbidge, C.A., Boss, P.K., Davies, C., 2013. Interactions between ethylene and auxin are crucial to the control of grape (Vitis vinifera L.) berry ripening. BMC Plant Biology 13(222), 1-14.

Boulton, R.B., Singleton, V.L., Bisson, L.F., Kunkee, R.E., 1999. Principles and Practices of Winemaking. Springer, US.

Butzke, C.E., 1998. Survey of Yeast Assimilable Nitrogen Status in Musts from California, Oregon, and Washington. Am. J. Enol. Vitic. 49(2), 220-224.

Camara, J., Alves, M., Marques, J., 2006. Multivariate analysis for the classification and differentiation of Madeira wines according to main grape varieties. Talanta 68(5), 1512–1521.

Carrau, F.M., Medina, K., Boido, E., Farina, L., Gaggero, C., Dellacassa, E., Versini, G., Henschke, P.A., 2005. De novo synthesis of monoterpenes by Saccharomyces cerevisiae wine yeasts FEMS Microbiol Lett 243(1), 107–115.

Comitini, F., Capece, A., Ciani, M., Romano, P., 2017. New insights on the use of wine yeasts. Current Opinion in Food Science 13, 44–49.

Cooper, T.G., 1982. Nitrogen Metabolism in Saccharomyces cerevisiae. In: Strathern, J.N., Jones, E.W., Broach, J.R. (Eds.). The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp. 39-99.

Cramer, G.R., Van Sluyter, S.C., Hopper, D.W., Pascovici, D., Keighley, T., Haynes, P.A., 2013. Proteomic analysis indicates massive changes in metabolism prior to the inhibition of growth and photosynthesis of grapevine (Vitis viniferaL.) in response to water deficit. BMC Plant Biology 13(49), 1-22.

Garde-Cerdán, T., Lorenzo, C., Lara, J.F., Pardo, F., Ancín-Azpilicueta, C., Salinas, M.R., 2009. Study of the Evolution of Nitrogen Compounds during Grape Ripening. Application to Differentiate Grape Varieties and Cultivated Systems Journal of Agricultural and Food Chemistry 57(6), 2410–2419.

Geana, E.I., Popescu, R., Costinel, D., Dinca, O.R., Ionete, R.E., Stefanescu, I., Artem, V., Bala, C., 2016. Classification of red wines using suitable markers coupled with multivariate statistic analysis. Food Chemistry 192, 1015–1024.

Gobbi, M., Comitini, F., D'Ignazi, G., Ciani, M., 2013. Effects of nutrient supplementation on fermentation kinetics, H2S evolution, and aroma profile in Verdicchio DOC wine production. European Food Research and Technology 236(1), 145–154.

Gobert, A., Tourdot-Maréchal, R., Morge, C., Sparrow, C., Liu, Y., Quintanilla-Casas, B., Vichi, S., Alexandre, H., 2017. Non-*Saccharomyces* Yeasts Nitrogen Source Preferences: Impact on Sequential Fermentation and Wine Volatile Compounds Profile. Front Microbiol 8, 1-13.

Guo, X.-W., Li, Y.-Z., Guo, J., Wang, Q., Huang, S.-Y., Chen, Y.-F., Du, L.-P., Xiao, D.-G., 2016. Reduced production of ethyl carbamate for wine fermentation by deleting *CAR1* in *Saccharomyces cerevisiae.* J. Ind. Microbiol. Biotechnol. 43(5), 671–679.

Hagen, K.M., Keller, M., Edwards, C.G., 2008. Survey of Biotin, Pantothenic acid, and Assimilable Nitrogen in Winegrapes from the Pacific Northwest. Am. J. Enol. Vitic. 59(4), 432–436.

Hannam, K.D., Neilsen, G.H., Neilsen, D., Midwood, A.J., Millard, P., Zhang, Z., Thornton, B., Steinke, D., 2016. Amino Acid Composition of Grape (Vitis vinifera L.) Juice in Response to Applications of Urea to the Soil or Foliage. American Journal of Enology and Viticulture 67(1), 47–55.

Henschke, P.A., Jiranek, V., 1993. Yeasts-metabolism of nitrogen compounds in Wine Microbiology and Biotechnology. Harwood Academic Publishers 77–164.

Hernández-Orte, P., Cacho, J.F., Ferreira, V., 2002. Relationship between Varietal Amino Acid Profile of Grapes and Wine Aromatic Composition. Experiments with Model Solutions and Chemometric Study. Journal of Agricultural and Food Chemistry 50(10), 2891–2899.

Huang, Z. & Ough, C.S., 1991. Amino Acid Profiles of Commercial Grape Juices and Wines. Am. J. Enol. Vitic. 42(3), 261–267.

Jolly, N.P., Augustyn, O.P.H., Pretorius, I.S., 2017. The Role and Use of Non-*Saccharomyces* Yeasts in Wine Production. South African Journal of Enology & Viticulture 27(1), 15-39.

Kliewer, W.M., 1969. Free Amino Acids and Other Nitrogenous Substances of Table Grape Varieties. Journal of Food Science 34(3), 274–278.

Kliewer, W.M., 1970. Free Amino Acids and Other Nitrogenous Fractions in Wine Grapes. Journal of Food Science 35(1), 17–21.

Landete, J.M., Ferrer, S., Pardo, I., 2007. Biogenic amine production by lactic acid bacteria, acetic bacteria and yeast isolated from wine. Food Control 18(12), 1569–1574.

Liu, L., Cozzolino, D., Cynkar, W.U., Gishen, M., Colby, C.B., 2006. Geographic Classification of Spanish and Australian Tempranillo Red Wines by Visible and Near-Infrared Spectroscopy Combined with Multivariate Analysis Journal of Agricultural and Food Chemistry 54(18), 6754–6759.

Loveys, B.R., Dry, P.R., Stoll, M., Mc Carthy, M.G., 2000. Using Plant Physiology to Improve the Water Use Efficiency of Horticultural Crops. Acta Horticulturae 537, 187-197.

Majumdar, R., Minocha, R., Minocha, S.C., 2015. Ornithine: at the crossroads of multiple paths to amino acids and polyamines. In: J.P.F. D'Mello (ed). Amino acids in higher plants. CABI, Wallingford. pp 156–176.

Mccarthy, M.G., 1997. The effect of transient water deficit on berry development of cv. Shiraz (Vitis vinifera L.). Australian Journal of Grape and Wine Research 3(3), 2–8.

Myles, S., Boyko, A.R., Owens, C.L., Brown, P.J., Grassi, F., Aradhya, M.K., Prins, B., Reynolds, A., Chia, J.-M., Ware, D., Bustamante, C.D., Buckler, E.S., 2011. Genetic structure and domestication history of the grape. PNAS 108(9), 3530–3535.

Nicolini, G., Larcher, R., Versini, G., 2004. Status of yeast assimilable nitrogen in Italian grape musts and effects of variety, ripening and vintage. Vitis 43(2), 89–96.

Ough, C.S. & Bell, A.A., 1980. Effects of Nitrogen Fertilization of Grapevines on Amino Acid Metabolism and Higher-Alcohol Formation during Grape Juice Fermentation. Am. J. Enol. Vitic. 31(2), 122–123.

Ough, C.S., Crowell, E.A., E.A., Gutlove, B.R., 1988. Carbamyl Compound Reactions with Ethanol. Am. J. Enol. Vitic. 39(3), 239–242.

Ough, C.S., Crowell, E.A., et al., 1988. Formation of Ethyl Carbamate Precursors During Grape Juice (Chardonnay) Fermentation. I. Addition of Amino Acids, Urea, and Ammonia: Effects of Fortification on Intracellular and Extracellular Precursors Am J Enol Vitic. 39(3), 243–249.

Ough, C.S., Huang, Z., Mooney, L.A., 1991. Amino Acid Uptake by Four Commercial Yeasts at Two Different Temperatures of Growth and Fermentation: Effects on Urea Excretion and Reabsorption. Am. J. Enol. Vitic. 42(1), 26–40.

Rapp, A. & Versini, G., 1991. Influence of nitrogen compounds in grapes on aroma compounds of wines In: Developments in Food Science 37. Elsevier 1659–1694.

Rollero, S., Bloem, A., Ortiz-Julien, A., Camarasa, C., Divol, B., 2018. Fermentation performances and aroma production of non-conventional wine yeasts are influenced by nitrogen preferences. FEMS Yeast Research 18(5), 1-11.

Romero, P., Muñoz, R.G., Fernández-Fernández, J.I., del Amor, F.M., Martínez-Cutillas, A., García-García, J., 2015. Improvement of yield and grape and wine composition in field-grown Monastrell grapevines by partial root zone irrigation, in comparison with regulated deficit irrigation. Agricultural Water Management 149, 55–73.

Saloua, H., Charlotte, G., Trotel-Aziz, P., Sandra, V., Rabenoelina, F., fabienne, B., Philippe, E., Clément, C., Ferchichi, A., Aziz, A., 2014. Drought stress tolerance in grapevine involves activation of polyamine oxidation contributing to improved immune response and low susceptibility to Botrytis cinerea. Journal of Experimental Botany 66, 1-13.

Serrano-Lourido, D., Saurina, J., Hernández-Cassou, S., Checa, A., 2012. Classification and characterisation of Spanish red wines according to their appellation of origin based on chromatographic profiles and chemometric data analysis. Food Chemistry 135(3), 1425–1431.

Smit, A.Y., 2013. The impact of nutrients on aroma and flavour production during wine fermentation. Thesis, Stellenbosch University, Private Bag X1, 7602 Matieland (Stellenbosch), South Africa.

Smit, A.Y., Engelbrecht, L., du Toit, M., 2012. Managing Your Wine Fermentation to Reduce the Risk of Biogenic Amine Formation. Front Microbiol 3, 1-10.

Soufleros, E.., Bouloumpasi, E., Tsarchopoulos, C., Biliaderis, C., 2003. Primary amino acid profiles of Greek white wines and their use in classification according to variety, origin and vintage. Food Chemistry 80(2), 261–273.

Spayd, S.E. & Andersen-Bagge, J., 1996. Free Amino Acid Composition of Grape Juice From 12 Vitis vinifera Cultivars in Washington. Am. J. Enol. Vitic. 47(4), 389–402.

Stines, A.P., Naylor, D.J., Høj, P.B., van Heeswijck, R., 1999. Proline Accumulation in Developing Grapevine Fruit Occurs Independently of Changes in the Levels of Δ1-Pyrroline-5-Carboxylate Synthetase mRNA or Protein. Plant Physiol 120(3), 923-931.

Stines, A.P., Grubb, J., Gockowiak, H., Henschke, P.A., HøJ, P.B., Heeswijck, R., 2000. Proline and arginine accumulation in developing berries of Vitis vinifera L. in Australian vineyards: Influence of vine cultivar, berry maturity and tissue type Australian Journal of Grape and Wine Research 6(2), 150–158.

Styger, G., Prior, B., Bauer, F.F., 2011. Wine flavor and aroma. J. Ind. Microbiol. Biotechnol. 38(9), 1145–1159.

Swiegers, J.H. & Pretorius, I.S., 2007. Modulation of volatile sulfur compounds by wine yeast. Appl Microbiol Biotechnol 74(5), 954–960.

Torrea, D., 2003. Production of volatile compounds in the fermentation of chardonnay musts inoculated with two strains of *Saccharomyces cerevisiae* with different nitrogen demands. Food Control 14(8), 565–571.

Ugliano, M., Henschke, P.A., Herderich, M.J., Pretorius, I.S., 2007. Nitrogen management is critical for wine flavour and style. Wine Industry Journal. 22(6), 24-30.

Vilanova, M., Ugliano, M., Varela, C., Siebert, T., Pretorius, I.S., Henschke, P.A., 2007. Assimilable nitrogen utilisation and production of volatile and non-volatile compounds in chemically defined medium by *Saccharomyces cerevisiae* wine yeasts. Applied Microbiology and Biotechnology 77(1), 145–157.

de Villiers, A., Majek, P., Lynen, F., Crouch, A., Lauer, H., Sandra, P., 2005. Classification of South African red and white wines according to grape variety based on the non-coloured phenolic content. European Food Research and Technology 221, 520–528.

Wang, S.S. & Brandriss, M.C., 1987. Proline utilization in *Saccharomyces cerevisiae*: sequence, regulation, and mitochondrial localization of the *PUT1* gene product. Mol Cell Biol 7(12), 4431–4440.

Appendix B

Chapter 5 Additional Tables and Figures

Chapter 5 Appendix B: Additional Tables and Figures

Table B5.1. Percentage of amino acids per cultivar (%).

Cultivar	Cabernet Franc	Cabernet Sauvignon	Chardon-nay	Chenin Blanc	Cinsaut	Grenache Blanc	Merlot	Pinotage	Rouss-anne	Sauvignon Blanc	Semillon	Shiraz	Viognier	Average
ALA	2.17	2.72	6.91	7.11	6.28	5.10	1.71	5.13	3.03	7.14	5.97	2.80	5.17	4,71
ARG	15.88	9.57	10.58	21.50	29.51	34.82	7.84	30.42	16.84	24.55	30.34	17.61	30.65	21,55
ASP	0.67	0.66	1.57	2.24	1.10	1.75	0.55	1.49	1.47	2.65	2.21	1.18	1.43	1,46
GABA	4.79	6.20	5.06	6.51	6.33	5.36	3.98	4.98	4.31	6.54	6.06	5.75	4.87	5,44
GLN	5.25	3.02	6.05	5.78	8.16	9.95	2.58	8.60	4.67	7.23	6.58	5.16	6.38	6,11
GLU	2.62	1.95	2.75	4.34	3.81	3.87	1.49	4.01	4.09	5.69	3.64	3.02	2.66	3,38
GLY	0.21	0.17	0.22	0.18	0.24	0.11	0.10	0.19	0.06	0.20	0.23	0.21	0.19	0,18
HIS	1.19	1.26	2.16	1.78	1.04	2.43	1.04	2.06	1.70	1.97	2.01	1.53	1.99	1,70
HYP	0.64	0.59	0.70	0.73	0.39	0.64	0.53	0.11	0.86	0.51	0.74	0.71	0.66	0,60
ILE	1.31	0.73	0.76	1.24	1.35	1.19	0.68	0.80	1.67	1.04	1.04	0.92	1.43	1,09
LEU	1.06	0.78	0.39	0.64	1.48	0.42	0.80	1.13	1.13	0.57	0.72	1.10	0.56	0,83
LYS	0.15	0.28	0.14	0.16	0.26	0.30	0.20	0.24	0.12	0.22	0.26	0.22	0.19	0,21
MET	0.07	0.15	0.47	0.32	0.21	0.08	0.03	0.36	0.09	0.19	0.48	0.05	0.07	0,20
ORN	0.10	0.05	0.02	0.08	0.22	0.34	0.06	0.15	0.02	0.11	0.21	0.03	0.12	0,12
PHE	0.61	0.88	1.19	2.39	1.68	3.07	0.43	0.91	2.53	2.14	1.52	0.84	2.37	1,58
PRO	49.59	61.79	46.36	27.34	20.71	14.42	69.21	20.79	37.10	24.56	22.08	44.56	25.95	35,73
SER	2.77	2.65	4.85	5.15	3.70	4.65	2.42	4.15	4.35	4.77	5.20	4.19	4.88	4,13
THR	2.03	2.06	3.95	6.72	4.26	3.92	1.50	4.61	3.37	4.98	4.76	3.33	5.08	3,89
TRP	7.54	3.25	4.45	3.77	7.52	6.01	3.76	8.39	10.89	3.32	4.09	5.42	3.43	5,53
VAL	1.36	1.23	1.41	2.02	1.75	1.59	1.09	1.48	1.72	1.63	1.85	1.37	1.91	1,57

Table B5.2. Mean concentrations of the various cultivars included in the survey (mg/L).

Cultivar	Cabernet Franc	Cabernet Sauvignon	Chardon-nay	Chenin Blanc	Cinsaut	Grenache Blanc	Merlot	Pinotage	Roussanne	Sauvignon Blanc	Semillon	Shiraz	Viognier
ALA	38.55	58.75	145.08	104.16	135.57	65.80	40.59	129.22	59.64	101.88	79.89	52.04	96.09
ARG	282.13	206.47	222.11	314.98	637.03	449.62	185.63	765.56	331.56	350.44	405.98	327.09	569.91
ASP	11.98	14.16	32.98	32.83	23.83	22.59	13.11	37.54	28.98	37.86	29.60	21.86	26.68
GABA	85.08	133.81	106.33	95.43	136.60	69.21	94.22	125.26	84.79	93.35	81.02	106.80	90.47
GLN	93.38	65.24	127.15	84.66	176.25	128.49	61.06	216.37	92.01	103.19	88.10	95.88	118.63
GLU	46.50	42.05	57.84	63.61	82.28	49.94	35.23	100.96	80.58	81.27	48.76	56.13	49.46
GLY	3.70	3.69	4.54	2.57	5.25	1.39	2.32	4.75	1.08	2.82	3.10	3.87	3.55
HIS	21.11	27.19	45.41	26.02	22.52	31.34	24.68	51.92	33.41	28.10	26.86	28.32	37.08
HYP	11.34	12.75	14.72	10.66	8.39	8.31	12.55	2.79	16.91	7.22	9.93	13.26	12.24
ILE	23.27	15.71	15.92	18.16	29.24	15.37	16.02	20.09	32.78	14.81	13.93	17.05	26.57
LEU	18.90	16.88	8.19	9.41	32.02	5.38	19.00	28.41	22.19	8.17	9.57	20.41	10.36
LYS	2.60	6.07	2.86	2.41	5.56	3.88	4.79	5.98	2.32	3.21	3.53	4.12	3.57
MET	1.19	3.28	9.96	4.62	4.57	1.06	0.66	8.99	1.68	2.69	6.43	0.88	1.27
ORN	1.83	1.12	0.47	1.10	4.75	4.40	1.31	3.69	0.43	1.51	2.80	0.60	2.15
PHE	10.93	19.03	25.06	35.00	36.31	39.66	10.09	22.89	49.75	30.48	20.28	15.53	44.09
PRO	881.33	1333.56	973.52	400.57	447.03	186.18	1638.2	523.28	730.35	350.55	295.45	827.50	482.41
SER	49.14	57.29	101.86	75.49	79.86	60.04	57.17	104.56	85.66	68.04	69.54	77.77	90.70
THR	36.05	44.48	83.01	98.48	91.91	50.60	35.49	116.10	66.27	71.13	63.65	61.86	94.49
TRP	133.96	70.08	93.42	55.27	162.33	77.66	89.02	211.11	214.29	47.33	54.75	100.58	63.86
VAL	24.19	26.65	29.66	29.58	37.74	20.52	25.91	37.22	33.85	23.33	24.74	25.53	35.55

Table B5.3. Standard deviations of the various cultivars included in the survey (mg/L).

Cultivar	Cabernet Franc	Cabernet Sauvignon	Chardon-nay	Chenin Blanc	Cinsaut	Grenache Blanc	Merlot	Pinotage	Rouss-anne	Sauvignon Blanc	Semillon	Shiraz	Viognier
ALA	18.33	48.50	78.35	58.60	67.06	30.25	24.46	53.34	70.85	59.43	45.87	35.37	43.41
ARG	224.02	168.24	134.43	201.49	321.92	288.64	233.83	412.63	314.71	249.18	219.29	212.79	278.74
ASP	10.88	16.47	26.66	24.87	14.27	15.02	10.49	12.91	26.91	28.02	25.64	15.03	21.02
GABA	34.77	56.88	46.93	50.24	39.32	28.53	24.80	45.96	75.18	43.88	32.96	39.30	48.30
GLN	77.47	45.05	122.30	81.22	101.65	114.47	47.32	139.20	115.11	91.02	82.34	80.03	107.34
GLU	38.16	36.61	34.16	40.91	31.92	38.64	19.64	39.60	43.50	53.43	21.00	30.49	28.58
GLY	1.84	1.79	2.11	1.94	2.19	1.49	1.76	3.63	1.55	1.98	2.32	3.66	1.89
HIS	18.33	17.72	27.84	20.51	26.27	23.98	13.43	45.05	60.93	23.85	18.10	17.36	23.90
HYP	7.85	7.99	11.80	8.57	6.21	8.23	7.78	5.45	17.61	6.74	7.29	7.76	9.56
ILE	10.12	7.55	8.86	11.77	20.35	10.52	8.30	13.67	27.72	11.68	5.58	6.65	10.37
LEU	14.57	10.63	11.13	13.30	24.55	8.29	10.82	21.13	15.40	11.96	11.21	11.79	13.99
LYS	2.66	7.64	3.85	3.01	2.07	4.75	4.48	4.81	3.18	4.05	2.77	4.21	3.99
MET	1.56	8.58	10.29	4.92	4.34	2.42	1.15	8.72	3.29	3.48	6.85	1.94	2.08
ORN	1.87	1.37	0.82	1.94	4.11	5.29	1.67	3.88	0.92	1.96	2.40	1.24	2.02
PHE	11.02	15.82	13.83	25.31	24.09	20.25	9.17	12.74	68.50	25.62	12.01	10.46	24.22
PRO	673.70	913.16	594.82	272.24	188.69	138.10	743.15	209.83	904.33	317.29	105.50	593.68	579.76
SER	16.05	30.26	39.73	38.54	21.69	37.08	15.19	31.90	50.06	32.94	24.05	51.28	42.02
THR	17.93	24.66	41.54	56.93	39.64	33.47	16.23	49.48	81.87	37.48	33.28	28.77	43.64
TRP	79.10	89.58	112.59	72.13	77.26	85.36	105.97	151.86	149.17	62.62	56.92	52.34	78.45
VAL	9.11	7.47	14.30	17.09	20.37	13.43	5.60	18.95	34.22	13.33	10.01	8.68	13.15

Table B5.4. Median concentrations of the various cultivars included in the survey (mg/L).

Cultivar	Cabernet Franc	Cabernet Sauvignon	Chardon-nay	Chenin Blanc	Cinsaut	Grenache Blanc	Merlot	Pinotage	Rouss-anne	Sauvignon Blanc	Semillon	Shiraz	Viognier
ALA	33.19	43.38	134.52	94.62	116.76	56.98	29.15	124.33	44.20	91.63	67.21	47.33	88.59
ARG	201.12	158.52	188.81	264.41	525.59	343.01	103.31	721.02	248.22	283.71	370.60	290.71	515.98
ASP	11.78	8.61	25.03	27.01	21.83	16.23	11.92	38.25	18.23	31.56	20.47	18.95	22.08
GABA	72.61	121.15	99.14	84.49	133.59	60.47	88.04	105.91	59.41	84.19	75.47	111.52	74.75
GLN	69.05	51.97	109.35	63.09	137.17	74.42	49.22	181.10	55.55	70.04	72.05	69.08	74.33
GLU	37.23	30.36	48.88	54.42	72.91	33.93	29.67	105.66	72.35	66.23	46.63	50.36	37.86
GLY	3.34	3.98	4.71	2.42	5.61	1.29	2.66	4.49	ND	2.69	2.75	3.55	3.75
HIS	18.82	27.14	37.82	19.79	12.44	24.49	26.89	57.82	20.99	21.10	24.12	26.47	38.32
HYP	12.02	12.93	12.18	9.24	6.62	6.55	13.13	ND	14.06	6.01	8.62	12.85	13.11
ILE	24.29	15.91	13.69	15.25	21.61	13.17	14.57	14.68	25.20	11.62	14.54	16.44	27.04
LEU	20.67	19.43	0.77	0.21	22.45	ND	19.08	28.36	25.41	1.06	1.66	18.95	0.53
LYS	2.07	3.64	1.92	1.77	5.22	1.37	3.82	6.77	ND	2.21	3.56	3.50	2.57
MET	0.13	0.13	7.25	3.78	2.08	ND	0.13	7.68	0.13	1.69	5.07	0.13	0.13
ORN	2.00	ND	ND	0.28	2.96	3.09	ND	3.47	ND	0.82	2.45	ND	2.33
PHE	7.49	15.07	21.66	28.97	26.91	34.58	7.39	21.38	33.26	23.73	18.47	15.43	37.21
PRO	908.65	1169.09	797.46	320.50	438.28	138.78	1884.42	507.14	284.72	274.19	291.72	707.67	326.98
SER	48.23	44.95	94.36	68.41	73.70	54.76	53.96	107.89	68.07	60.72	73.98	65.17	79.26
THR	39.59	36.97	76.34	85.16	76.79	35.15	31.94	120.66	41.65	61.94	53.14	60.83	90.13
TRP	138.47	33.75	39.58	18.16	153.71	45.15	63.74	203.24	249.81	21.27	33.35	94.84	25.54
VAL	25.55	27.76	25.49	25.72	28.22	14.07	26.42	37.41	26.84	20.33	26.25	24.85	32.42

Table B5.5. Minimum concentrations of the various cultivars included in the survey (mg/L).

Cultivar	Cabernet Franc	Cabernet Sauvignon	Chardon-nay	Chenin Blanc	Cinsaut	Grenache Blanc	Merlot	Pinotage	Rouss-anne	Sauvignon Blanc	Semillon	Shiraz	Viognier
ALA	12.43	5.06	18.65	12.43	50.68	36.77	15.99	55.08	13.72	0.03	17.72	0.55	11.84
ARG	30.79	13.56	43.69	32.52	271.11	102.32	30.96	108.99	44.04	21.38	110.27	22.56	44.65
ASP	ND	ND	ND	ND	ND	4.26	ND	10.74	3.47	ND	ND	ND	3.06
GABA	44.79	54.77	30.53	14.47	78.34	40.28	50.54	75.15	38.90	13.90	37.43	ND	25.47
GLN	0.11	0.11	ND	ND	69.61	13.80	0.11	36.07	21.44	ND	ND	10.55	ND
GLU	ND	ND	12.46	13.75	44.53	16.71	ND	33.28	24.72	11.29	18.86	13.80	12.20
GLY	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
HIS	ND	ND	ND	ND	ND	ND	ND	ND	3.93	ND	ND	ND	ND
HYP	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
ILE	11.53	0.46	0.46	ND	11.56	0.46	0.46	0.46	11.12	ND	5.12	0.46	8.42
LEU	ND	ND	ND	ND	12.44	ND	ND	ND	ND	ND	ND	ND	ND
LYS	ND	ND	ND	ND	2.38	ND	ND	ND	ND	ND	ND	ND	ND
MET	0.13	0.13	ND	ND	0.13	ND	0.13	0.13	ND	ND	ND	0.13	ND
ORN	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
PHE	ND	ND	7.29	ND	13.43	15.78	ND	ND	7.42	ND	6.45	ND	11.88
PRO	178.96	76.37	130.11	78.43	235.50	33.22	159.70	257.60	195.74	62.53	111.94	51.18	141.55
SER	24.98	27.32	28.96	0.17	52.13	24.82	23.73	53.55	41.86	7.98	29.15	19.42	36.75
THR	ND	ND	12.38	12.03	43.39	18.55	14.31	40.41	20.08	7.31	21.19	11.61	9.87
TRP	ND	ND	6.24	ND	36.24	8.78	ND	46.40	7.96	ND	3.10	7.39	ND
VAL	12.06	6.98	ND	6.19	21.34	1.31	13.14	12.84	13.98	ND	10.59	7.89	16.18

Table B5.6. Maximum concentrations of the various cultivars included in the survey (mg/L).

Cultivar	Cabernet Franc	Cabernet Sauvignon	Chardon-nay	Chenin Blanc	Cinsaut	Grenache Blanc	Merlot	Pinotage	Rouss-anne	Sauvignon Blanc	Semillon	Shiraz	Viognier
ALA	73.59	215.38	538.30	433.21	254.52	141.10	99.81	232.81	311.35	465.02	211.39	176.98	234.83
ARG	712.47	771.80	775.51	1345.35	1251.89	1150.68	1210.58	1616.56	1381.61	1406.85	924.14	1067.62	1332.49
ASP	44.21	66.44	121.53	122.38	53.99	57.45	43.90	66.43	105.08	163.94	83.21	65.68	109.45
GABA	169.01	285.22	261.76	389.66	208.67	126.94	145.56	217.31	328.02	331.51	148.91	188.84	244.69
GLN	274.74	203.64	988.05	546.35	372.29	398.81	181.41	552.57	477.64	470.88	351.37	469.23	379.27
GLU	155.47	172.23	207.44	244.46	160.28	145.10	80.76	162.28	190.81	428.29	97.51	153.59	127.74
GLY	6.97	7.30	9.68	11.77	8.61	6.32	5.78	11.61	4.73	9.67	8.21	23.25	7.59
HIS	53.58	61.28	139.79	131.76	93.32	75.51	56.95	140.21	251.38	162.91	56.38	83.45	99.36
HYP	29.77	36.51	86.88	49.48	24.83	28.44	32.04	16.47	68.16	42.25	21.87	35.07	35.43
ILE	36.69	36.01	41.69	72.55	69.85	38.91	41.40	50.13	129.01	70.74	21.33	36.59	54.08
LEU	47.38	40.27	50.23	53.65	83.07	29.48	52.63	74.94	41.70	67.52	30.47	72.87	41.51
LYS	8.19	34.19	21.44	15.73	9.21	12.68	18.73	11.75	9.94	20.51	9.77	23.16	13.97
MET	4.91	39.60	43.73	28.08	12.52	8.53	4.40	24.68	8.99	22.59	19.00	10.11	7.27
ORN	4.86	4.19	3.18	12.63	12.85	19.47	5.13	10.45	3.31	10.71	8.68	6.29	7.08
PHE	26.58	81.75	77.81	195.12	88.61	79.15	32.96	47.80	293.21	165.30	51.87	53.75	118.05
PRO	2122.94	3171.60	3148.46	1810.55	1028.90	550.07	2637.99	919.36	2835.06	2593.47	455.05	3445.43	3283.74
SER	83.94	162.78	236.36	310.53	123.24	176.76	91.04	170.10	243.10	254.07	109.20	275.72	211.24
THR	63.79	108.69	240.60	373.20	158.13	138.70	83.38	201.04	355.98	238.12	152.04	158.67	214.78
TRP	262.46	431.74	583.73	331.87	298.91	298.20	513.04	556.15	426.27	386.13	222.26	245.40	286.53
VAL	4ND	40.76	81.79	91.32	78.68	43.64	38.32	82.83	154.41	82.83	41.96	49.50	74.47

Table B5.7. Range of amino acid concentrations of the various cultivars included in the survey (mg/L).

Cultivar	Cabernet Franc	Cabernet Sauvignon	Chardon-nay	Chenin Blanc	Cinsaut	Grenache Blanc	Merlot	Pinotage	Rouss-anne	Sauvignon Blanc	Semillon	Shiraz	Viognier
ALA	61.16	210.32	519.65	420.77	203.85	104.33	83.82	177.73	297.63	464.99	193.67	176.43	222.99
ARG	681.67	758.25	731.82	1312.83	980.78	1048.36	1179.62	1507.58	1337.56	1385.47	813.87	1045.06	1287.84
ASP	44.21	66.44	121.53	122.38	53.98	53.19	43.89	55.69	101.61	163.94	83.20	65.67	106.39
GABA	124.22	230.45	231.23	375.19	130.34	86.66	95.02	142.17	289.13	317.61	111.48	188.84	219.22
GLN	274.63	203.53	988.05	546.35	302.67	385.01	181.30	516.50	456.20	470.88	351.37	458.68	379.27
GLU	155.47	172.23	194.97	230.71	115.75	128.39	80.76	129.00	166.09	417.00	78.65	139.80	115.54
GLY	6.97	7.30	9.68	11.77	8.61	6.32	5.78	11.61	4.73	9.67	8.21	23.25	7.59
HIS	53.58	61.28	139.79	131.75	93.31	75.51	56.95	140.21	247.45	162.91	56.38	83.44	99.36
HYP	29.77	36.51	86.88	49.48	24.83	28.44	32.04	16.47	68.16	42.25	21.87	35.07	35.43
ILE	25.16	35.55	41.23	72.55	58.29	38.45	40.94	49.67	117.89	70.74	16.21	36.13	45.66
LEU	47.38	40.27	50.23	53.65	70.62	29.48	52.63	74.94	41.70	67.52	30.47	72.87	41.51
LYS	8.19	34.19	21.44	15.73	6.83	12.68	18.73	11.75	9.94	20.51	9.77	23.16	13.97
MET	4.78	39.46	43.73	28.08	12.39	8.53	4.27	24.55	8.99	22.59	19.00	9.98	7.27
ORN	4.86	4.19	3.18	12.63	12.85	19.47	5.13	10.45	3.31	10.71	8.68	6.29	7.08
PHE	26.58	81.75	70.52	195.12	75.17	63.38	32.96	47.80	285.79	165.30	45.42	53.75	106.17
PRO	1943.98	3095.23	3018.35	1732.12	793.40	516.85	2478.29	661.76	2639.32	2530.93	343.11	3394.25	3142.19
SER	58.96	135.46	207.40	310.36	71.11	151.94	67.31	116.54	201.24	246.09	80.05	256.30	174.50
THR	63.79	108.69	228.21	361.17	114.75	120.15	69.07	160.63	335.89	230.81	130.85	147.06	204.91
TRP	262.46	431.74	577.50	331.87	262.67	289.42	513.04	509.76	418.31	386.13	219.17	238.01	286.53
VAL	27.94	33.78	81.79	85.13	57.34	42.33	25.18	69.99	140.43	82.83	31.37	41.62	58.29

Table B5.8. Interquartile range of the various cultivars included in the survey (mg/L).

Cultivar	Cabernet Franc	Cabernet Sauvignon	Chardon-nay	Chenin Blanc	Cinsaut	Grenache Blanc	Merlot	Pinotage	Rouss-anne	Sauvignon Blanc	Semillon	Shiraz	Viognier
ALA	27.76	54.24	76.37	72.95	130.57	43.96	28.68	64.63	22.60	59.01	54.10	26.97	53.03
ARG	374.25	246.85	158.36	241.46	598.36	382.08	119.11	559.88	192.96	273.58	310.94	219.84	340.27
ASP	7.94	18.11	29.46	24.92	15.53	22.54	12.31	11.35	24.71	32.32	38.87	19.30	22.18
GABA	40.50	71.03	68.89	51.97	49.15	32.97	37.20	70.33	36.38	50.68	55.55	49.38	53.37
GLN	50.06	65.00	111.80	82.80	145.08	199.87	72.09	182.44	87.42	97.81	73.42	85.90	129.38
GLU	34.60	33.86	31.27	45.90	34.37	33.89	28.30	46.21	58.46	64.50	21.42	33.18	40.74
GLY	2.21	2.38	2.48	1.94	3.03	1.30	3.96	5.87	1.93	2.26	2.15	2.46	2.33
HIS	31.83	29.29	25.14	23.21	30.51	39.08	15.53	54.83	17.75	26.31	30.69	19.92	30.21
HYP	10.32	10.01	11.69	7.81	5.57	13.69	11.72	5.22	20.78	6.59	11.22	7.94	17.01
ILE	21.78	8.40	9.65	10.99	22.06	12.15	5.85	16.37	12.42	8.61	11.48	7.99	14.17
LEU	23.03	16.99	16.28	19.58	32.46	11.07	12.46	28.84	36.69	14.72	20.85	9.21	20.89
LYS	4.90	8.71	4.15	3.73	3.88	8.82	3.85	10.69	5.50	4.43	4.92	5.32	5.82
MET	2.22	2.82	15.19	6.63	7.51	0.26	ND	15.17	ND	4.29	12.88	ND	2.22
ORN	3.71	2.17	0.54	1.65	5.74	4.11	2.60	6.85	0.65	2.50	3.44	0.25	3.36
PHE	22.92	14.59	12.98	21.18	21.17	32.39	9.43	18.66	14.35	19.75	15.81	10.08	27.23
PRO	1259.06	1613.80	593.91	201.27	202.55	149.00	955.29	275.49	198.17	145.23	177.71	371.31	207.45
SER	18.05	34.52	47.59	41.41	32.59	38.48	22.07	34.21	56.98	36.93	38.09	50.23	42.86
THR	28.85	31.13	38.11	60.38	83.40	45.29	23.40	52.67	25.95	41.96	38.62	36.41	50.35
TRP	117.69	80.93	129.71	76.02	120.62	106.11	60.35	141.47	306.08	49.12	69.84	68.87	70.87
VAL	15.72	9.65	14.88	18.48	26.12	23.19	5.69	20.79	15.68	12.78	17.59	11.28	16.89

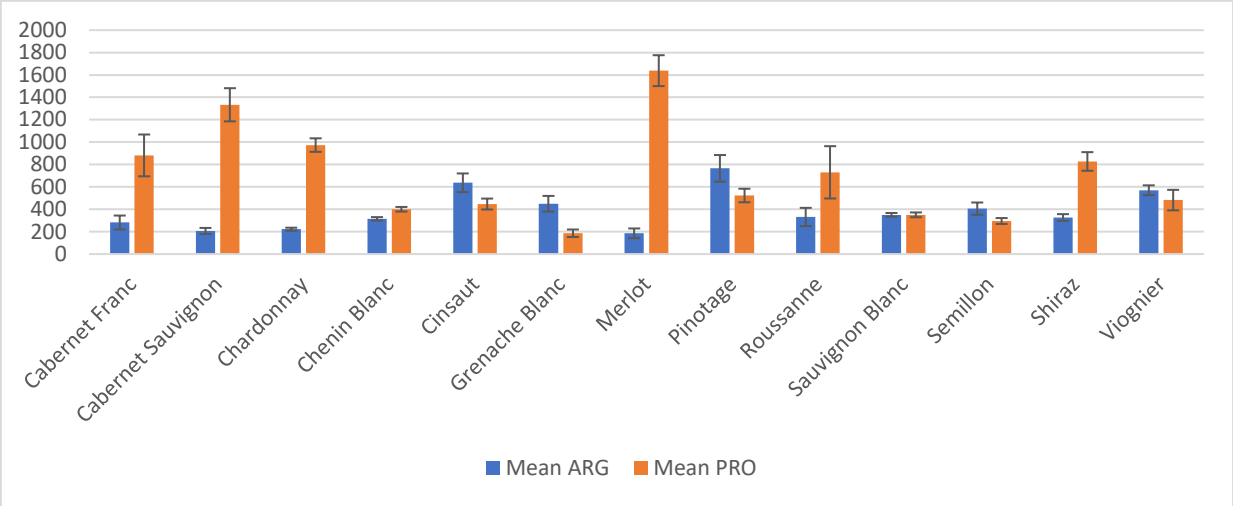


Figure B5.1. Mean proline and arginine concentrations of the various cultivars investigated in the survey (mg/L). Error bars indicate the standard error.

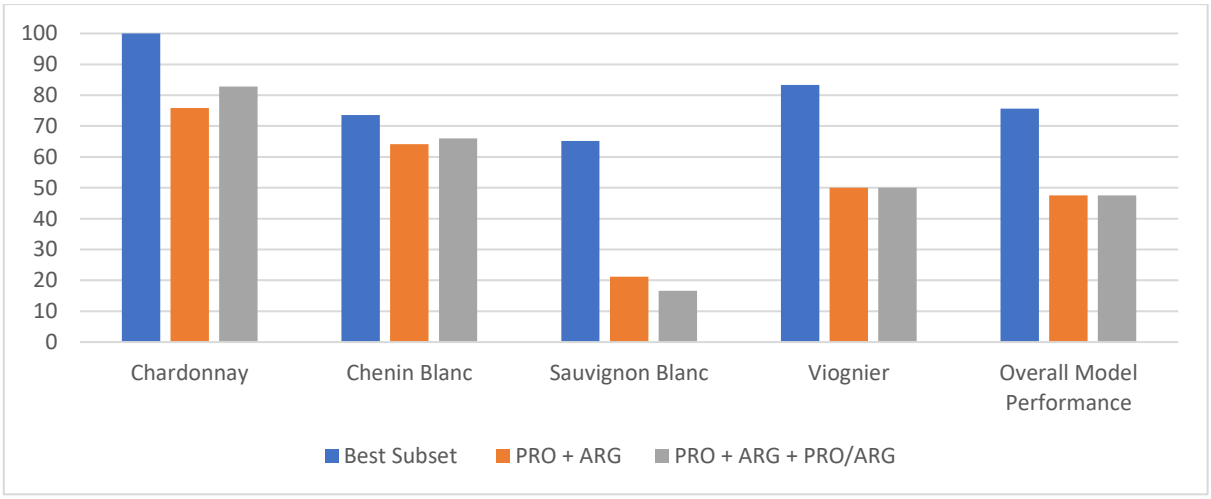


Figure B5.2. Overall percentage of white cultivars correctly predicted based on the various groups of predictor variables.

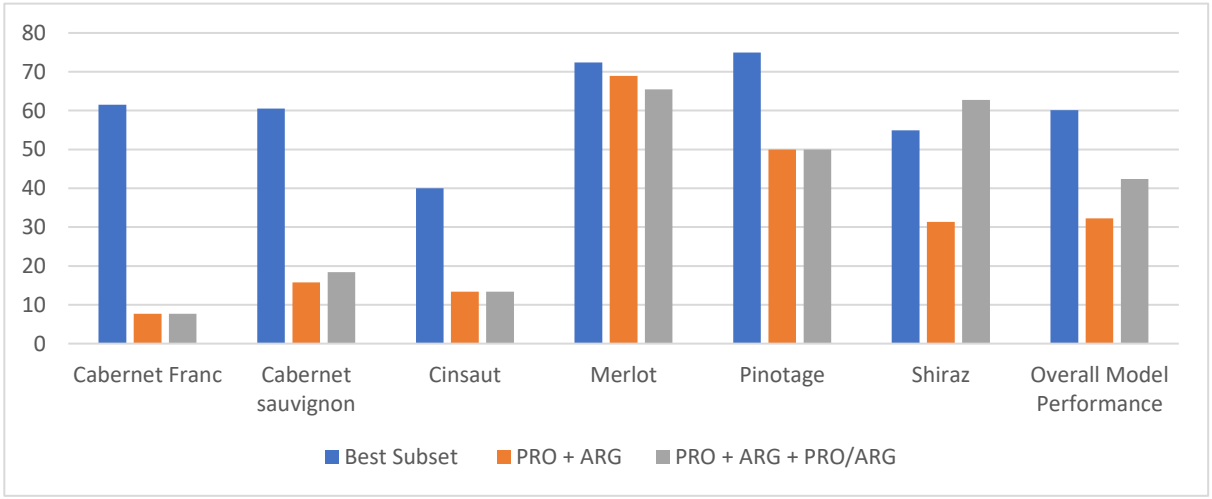


Figure B5.3. Overall percentage of red cultivars correctly predicted based on the various groups of predictor variables.

Table B5.9. Wilks Lambda test of significance of the various predictor variables used for the classification of red cultivars vs. white. *p*-values indicated in red are significant at α of 0.05.

Red vs. White	Wilks Lambda Test p-value
Best subset	
ALA	0,000000
GABA	0,000000
LEU	0,000000
PRO	0,000026

Table B5.10. Wilks Lambda test of significance of the various predictor variables used for the classification of white cultivars. *p*-values indicated in red are significant at α of 0.05.

White cultivars	Wilks Lambda Test p-value
Best subset	
ALA	0.000000
ARG	0.000000
GLU	0.000000
MET	0.000000
PRO	0.000000
THR	0.000000
PRO + ARG	
ARG	0.000000
PRO	0.000000
PRO + ARG + PRO/ARG	
ARG	0.000000
PRO	0.000000
PRO/ARG	0.000000

Table B5.11. Wilks Lambda test of significance of the various predictor variables used for the classification of red cultivars. *p*-values indicated in red are significant at α of 0.05.

Red cultivars	Wilks Lambda Test: *p-value*
Best subset	
GABA	0.000002
HYP	0.000040
ILE	0.000085
PHE	0.000004
PRO	0.000000
THR	0.000000
PRO + ARG	
ARG	0.000000
PRO	0.000224
PRO + ARG + PRO/ARG	
ARG	0.000000
PRO	0.017204
PRO/ARG	0.334248

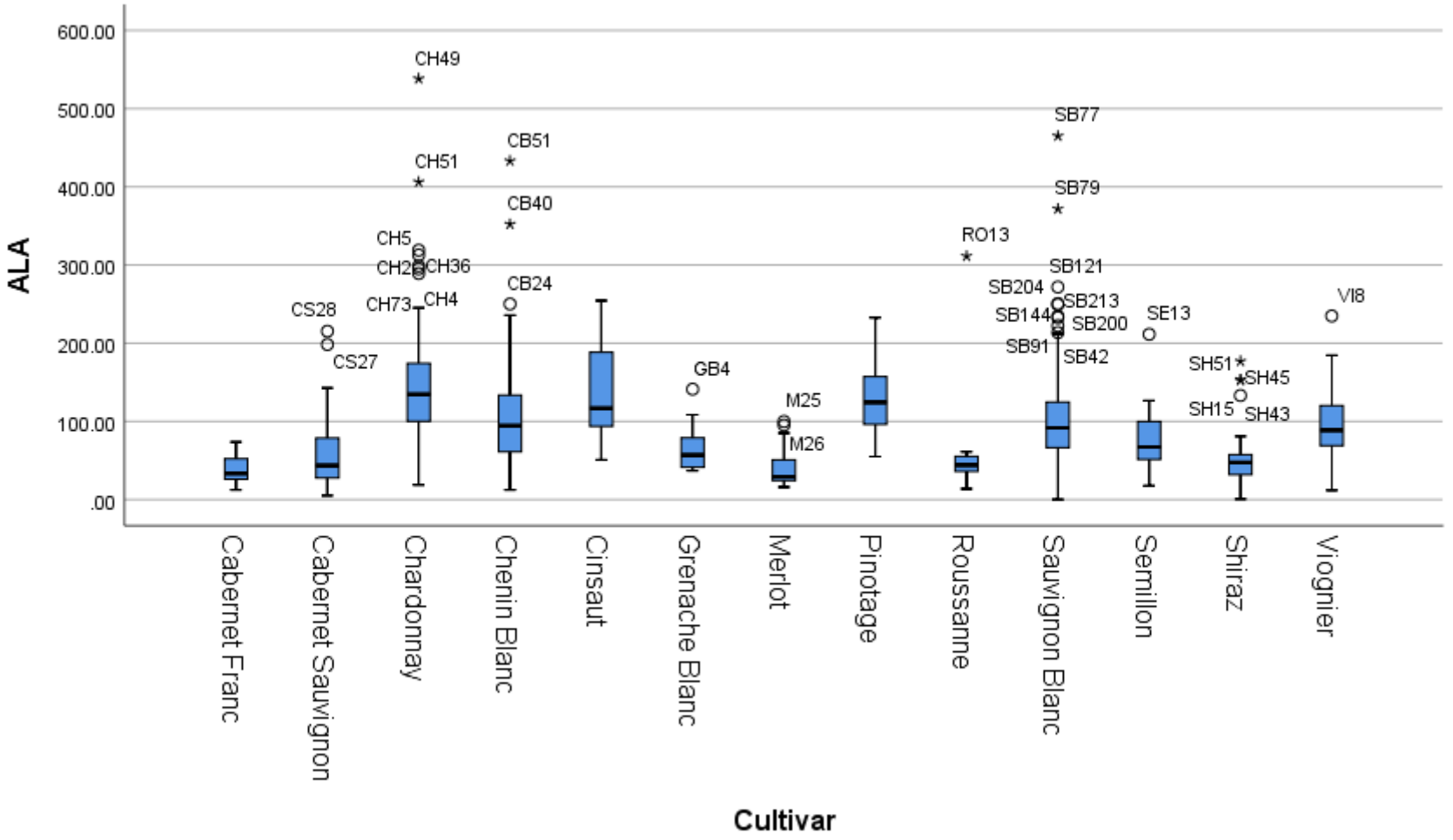


Figure B5.4. Box plots of alanine concentrations per cultivar (mg/L).

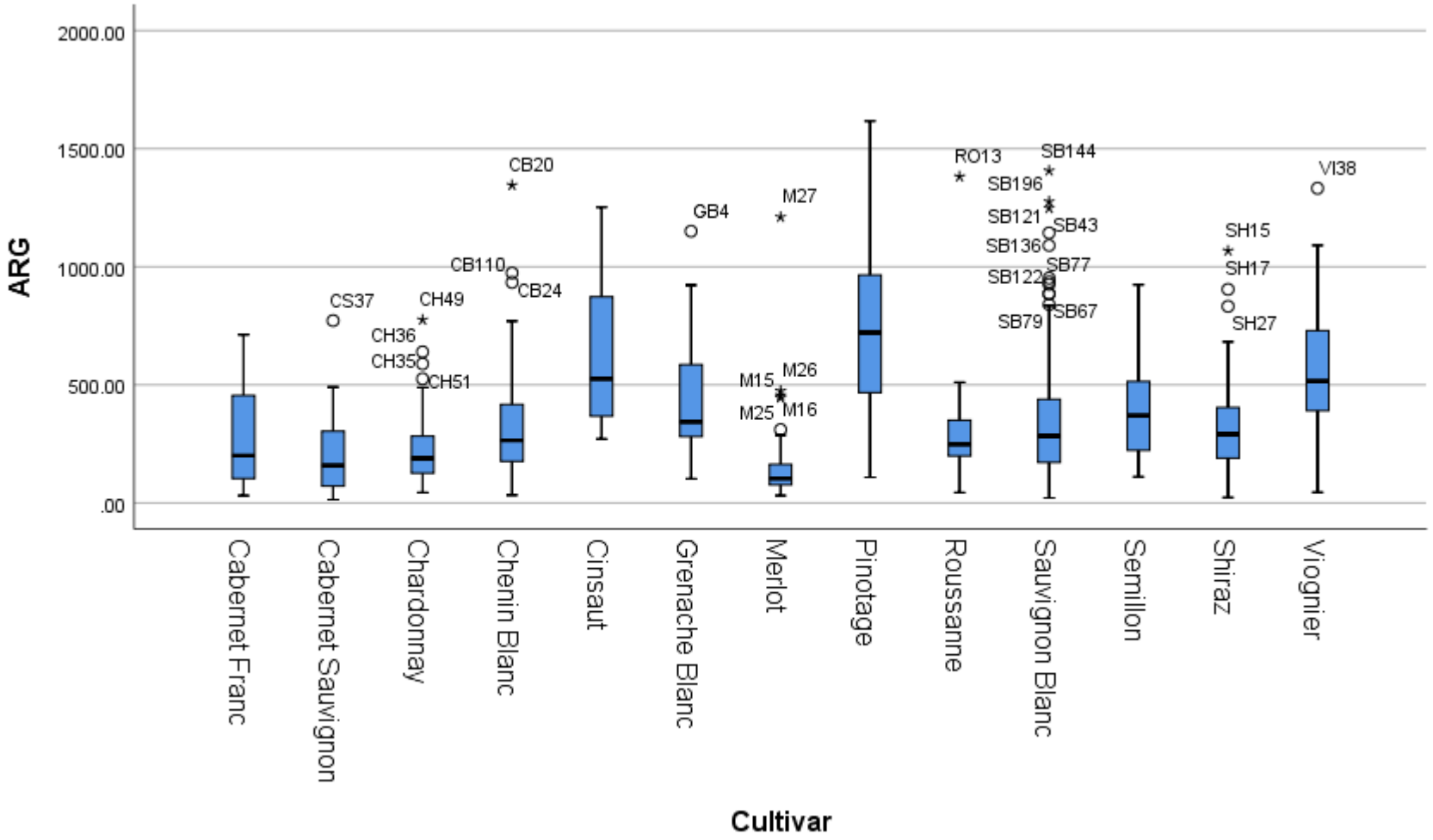


Figure B5.5. Box plots of arginine concentrations per cultivar (mg/L).

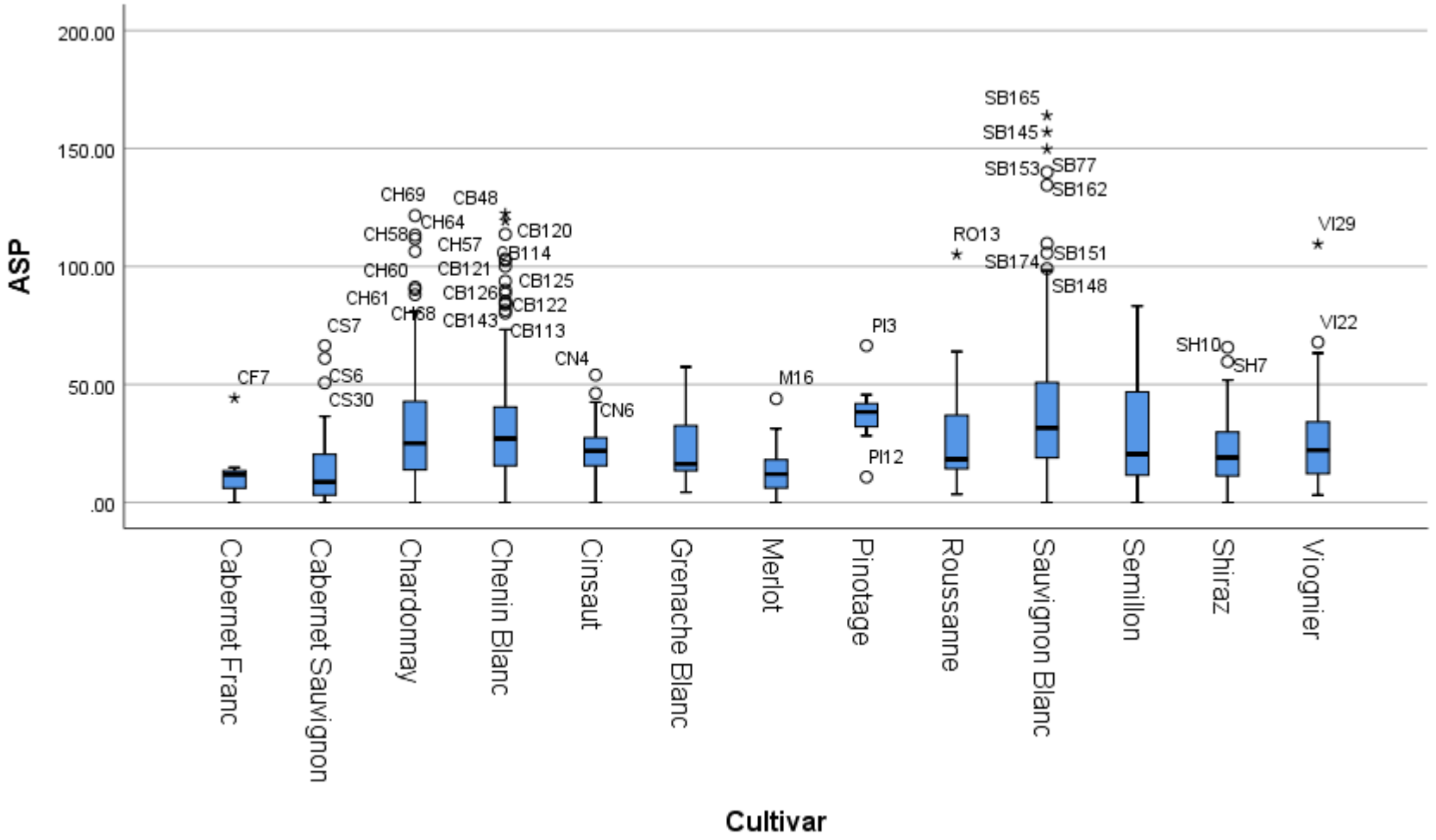


Figure B5.6. Box plots of aspartic acid concentrations per cultivar (mg/L).

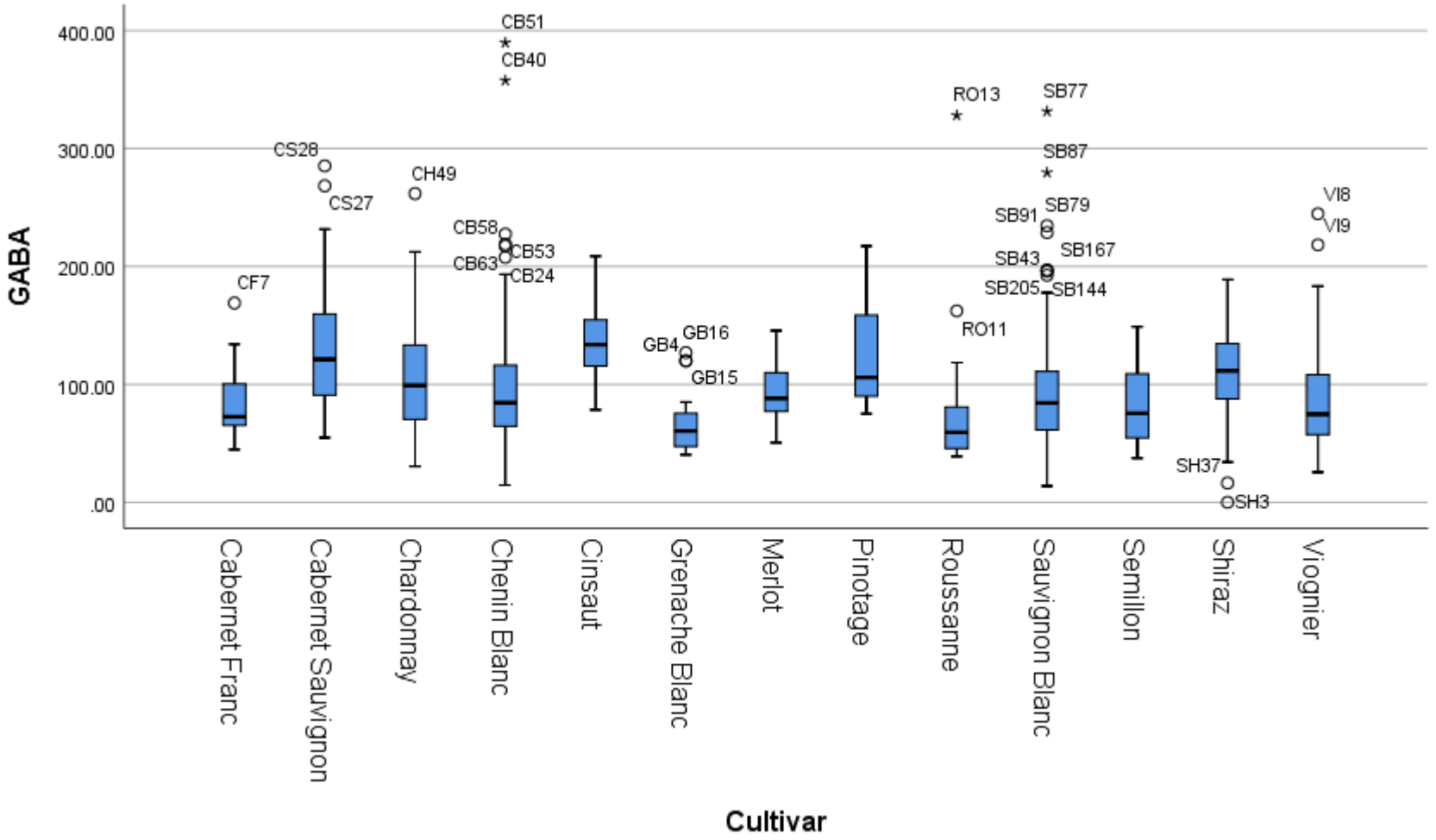


Figure B5.7. Box plots of GABA concentrations per cultivar (mg/L).

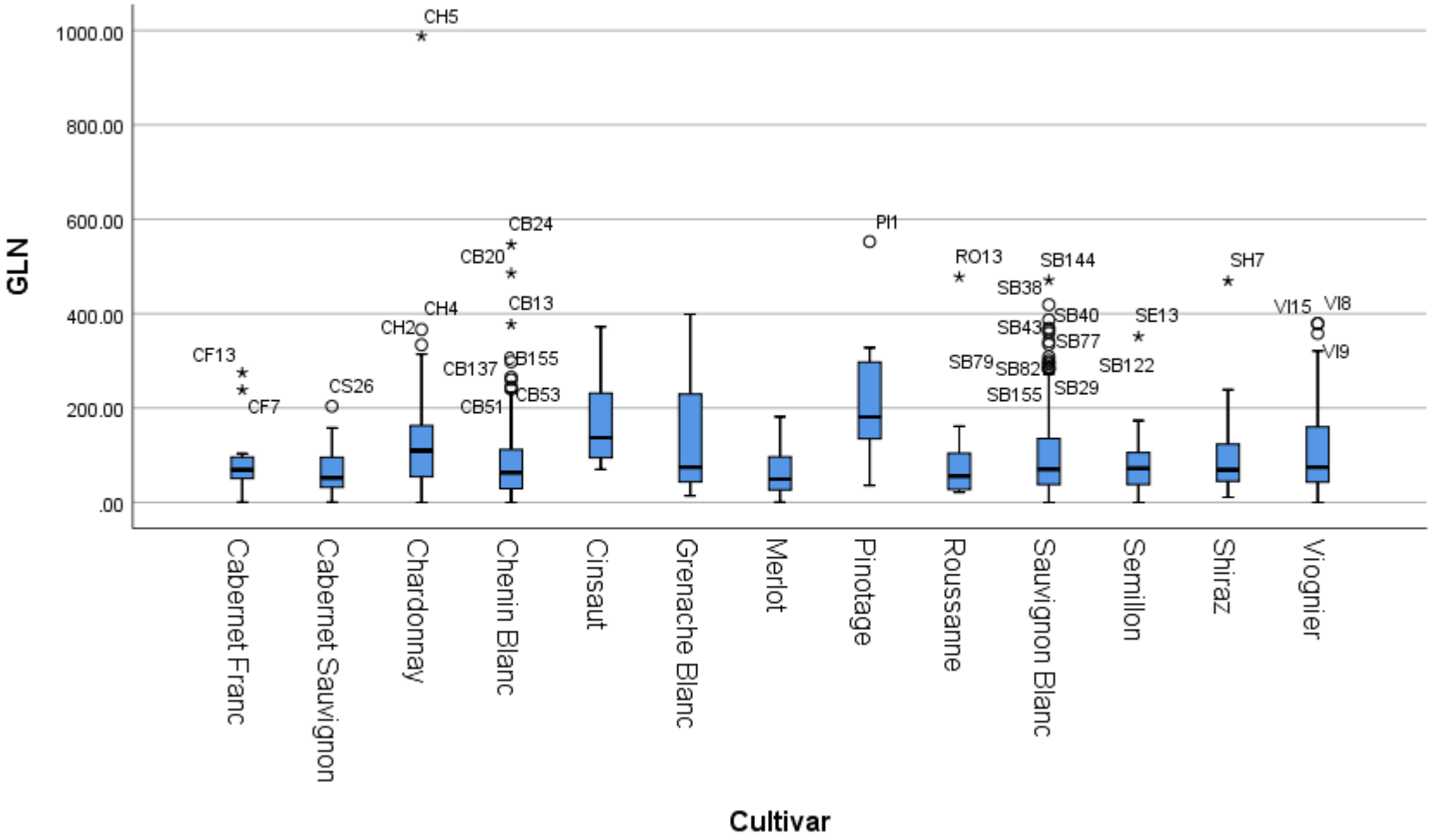


Figure B5.8. Box plots of glutamine concentrations per cultivar (mg/L).

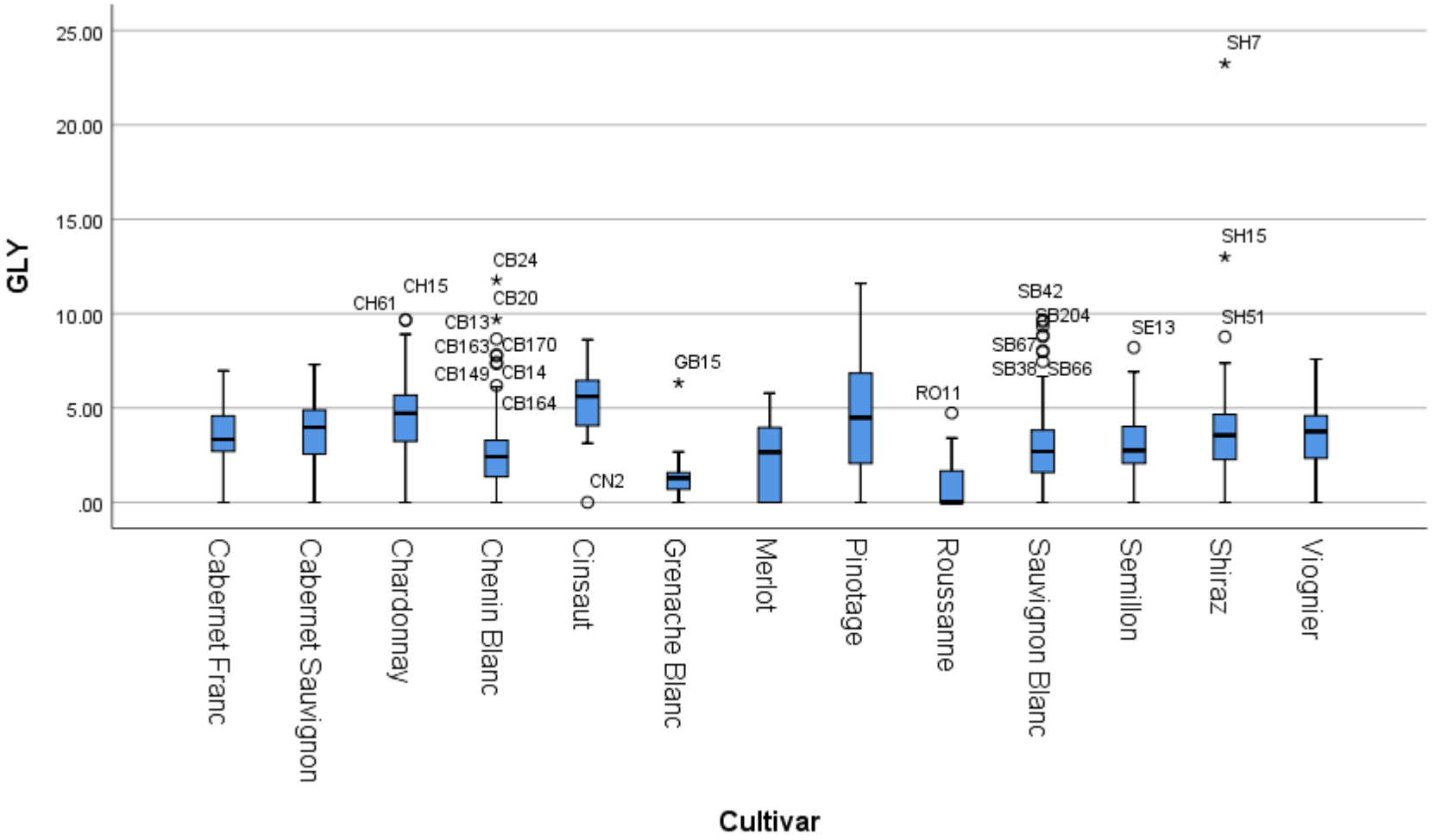


Figure B5.9. Box plots of glycine concentrations per cultivar (mg/L).

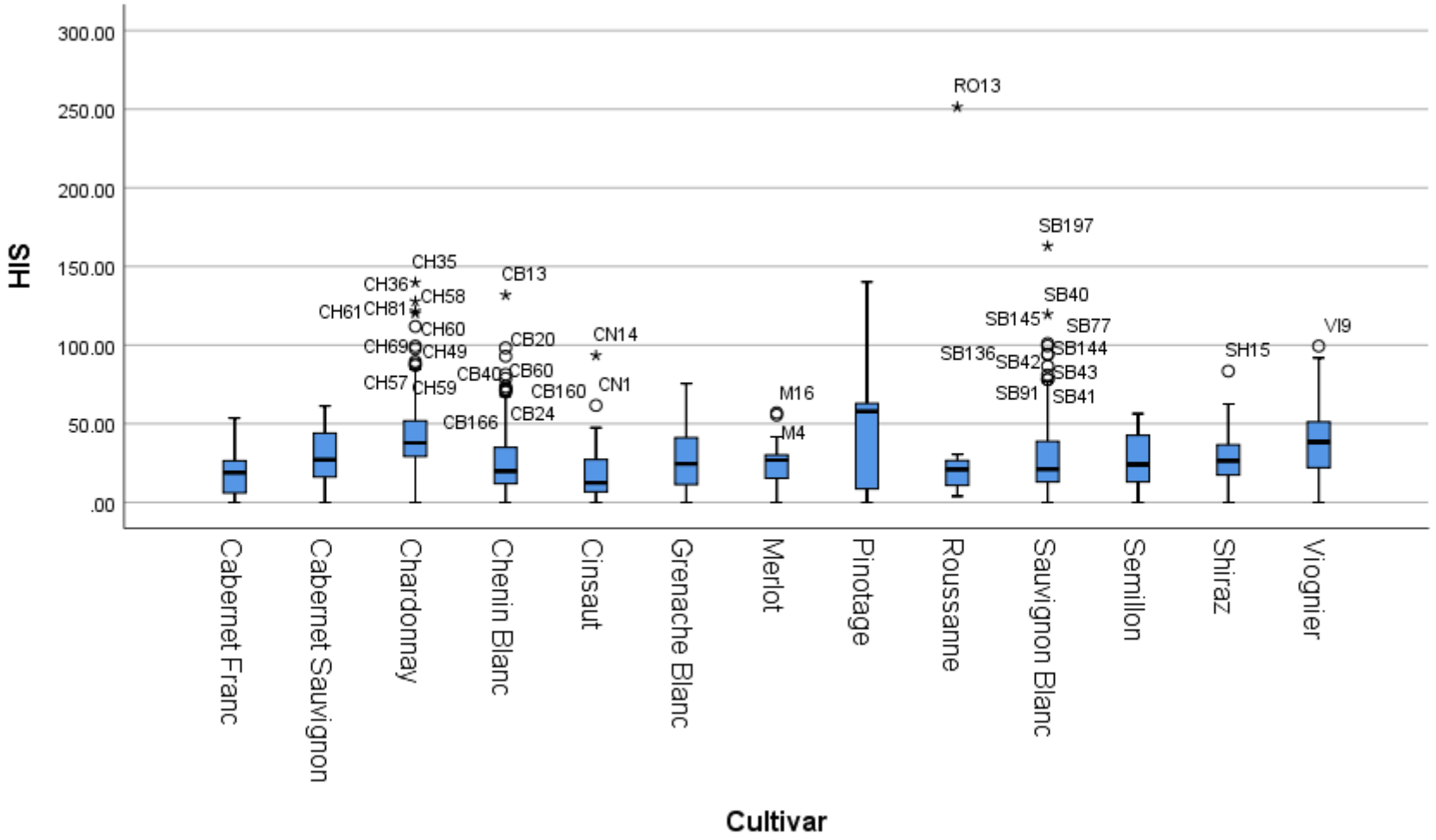


Figure B5.10. Box plots of histidine concentrations per cultivar (mg/L).

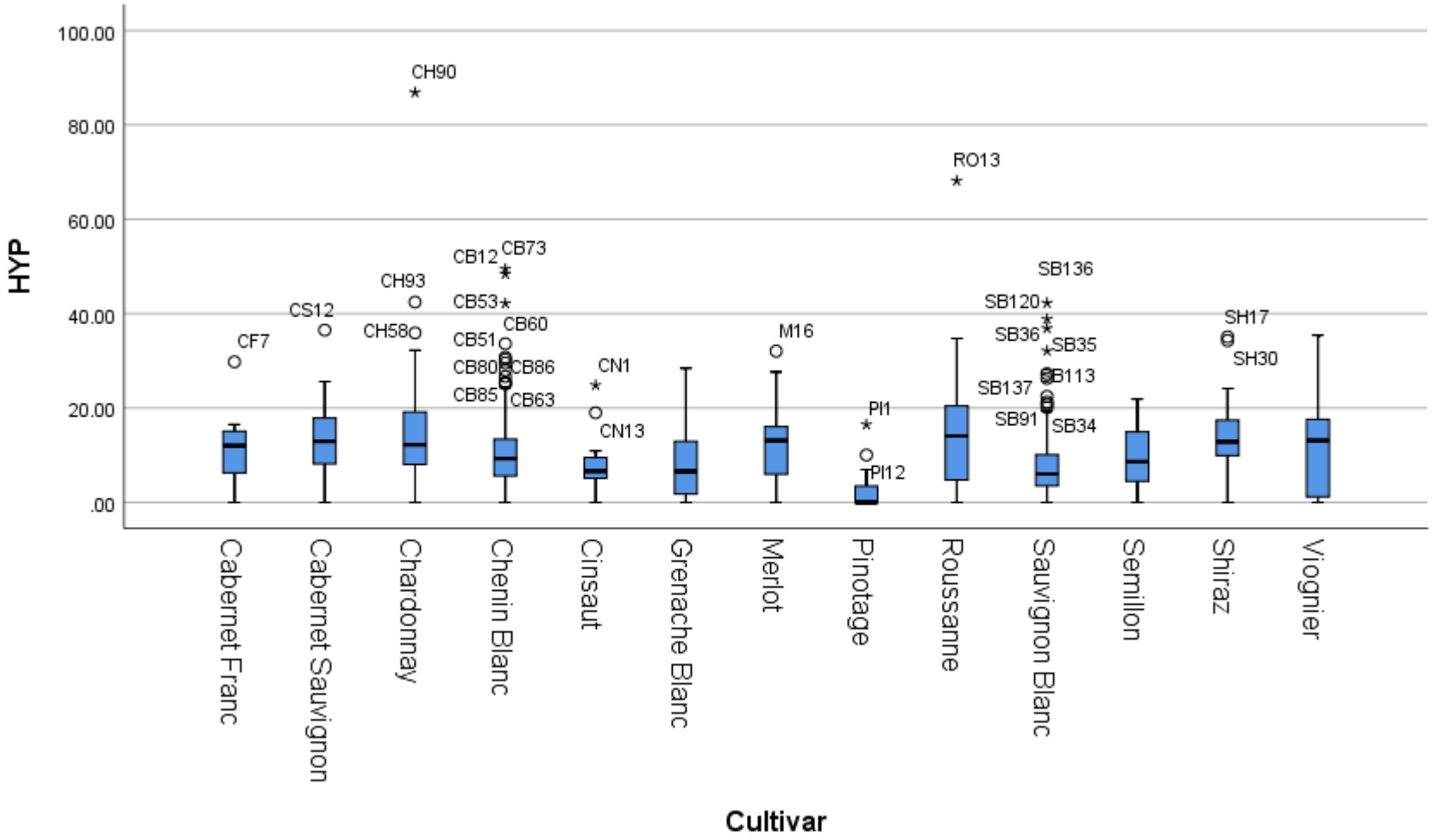


Figure B5.11. Box plots of hydroxyproline concentrations per cultivar (mg/L).

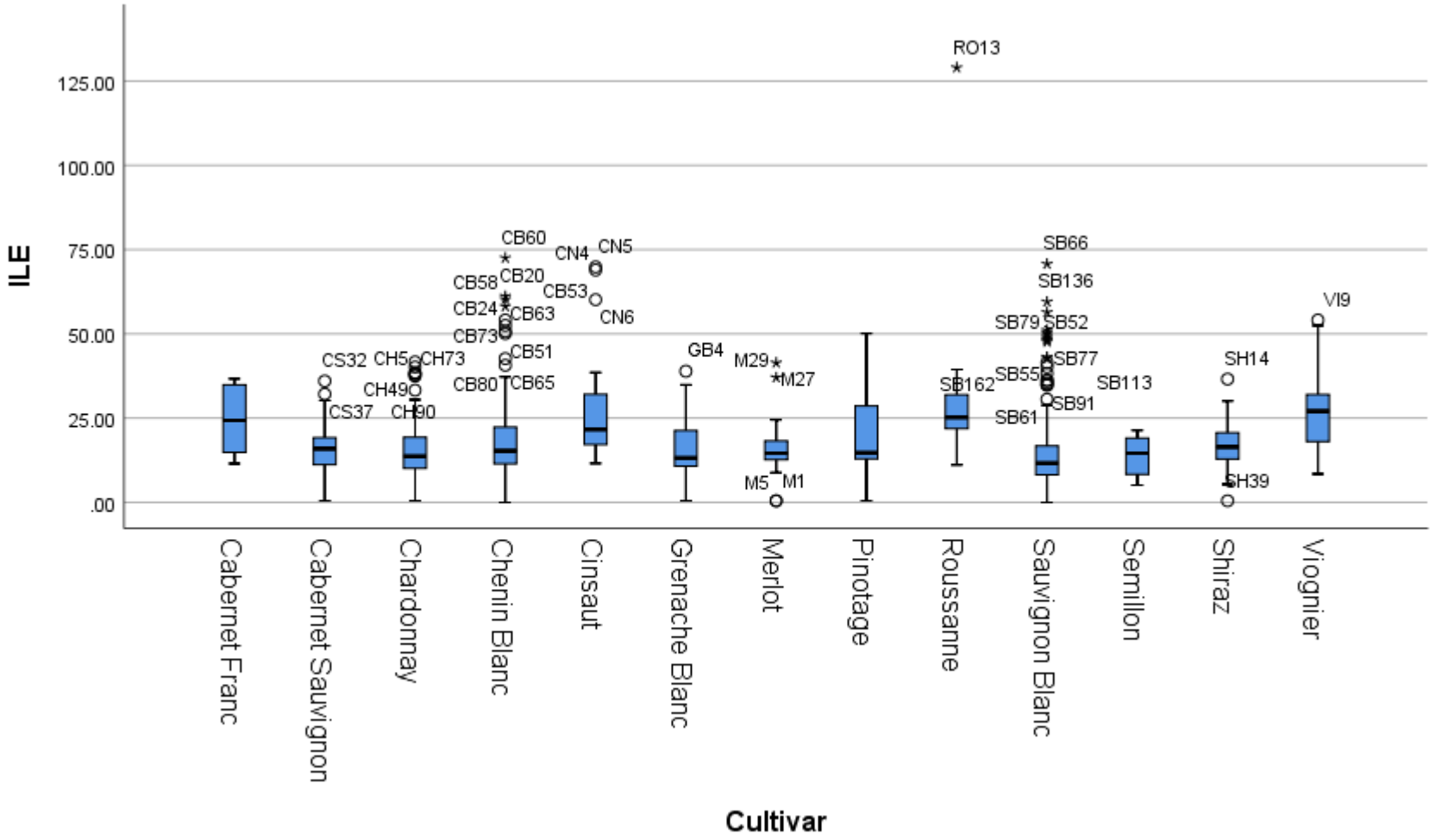


Figure B5.12. Box plots of isoleucine concentrations per cultivar (mg/L).

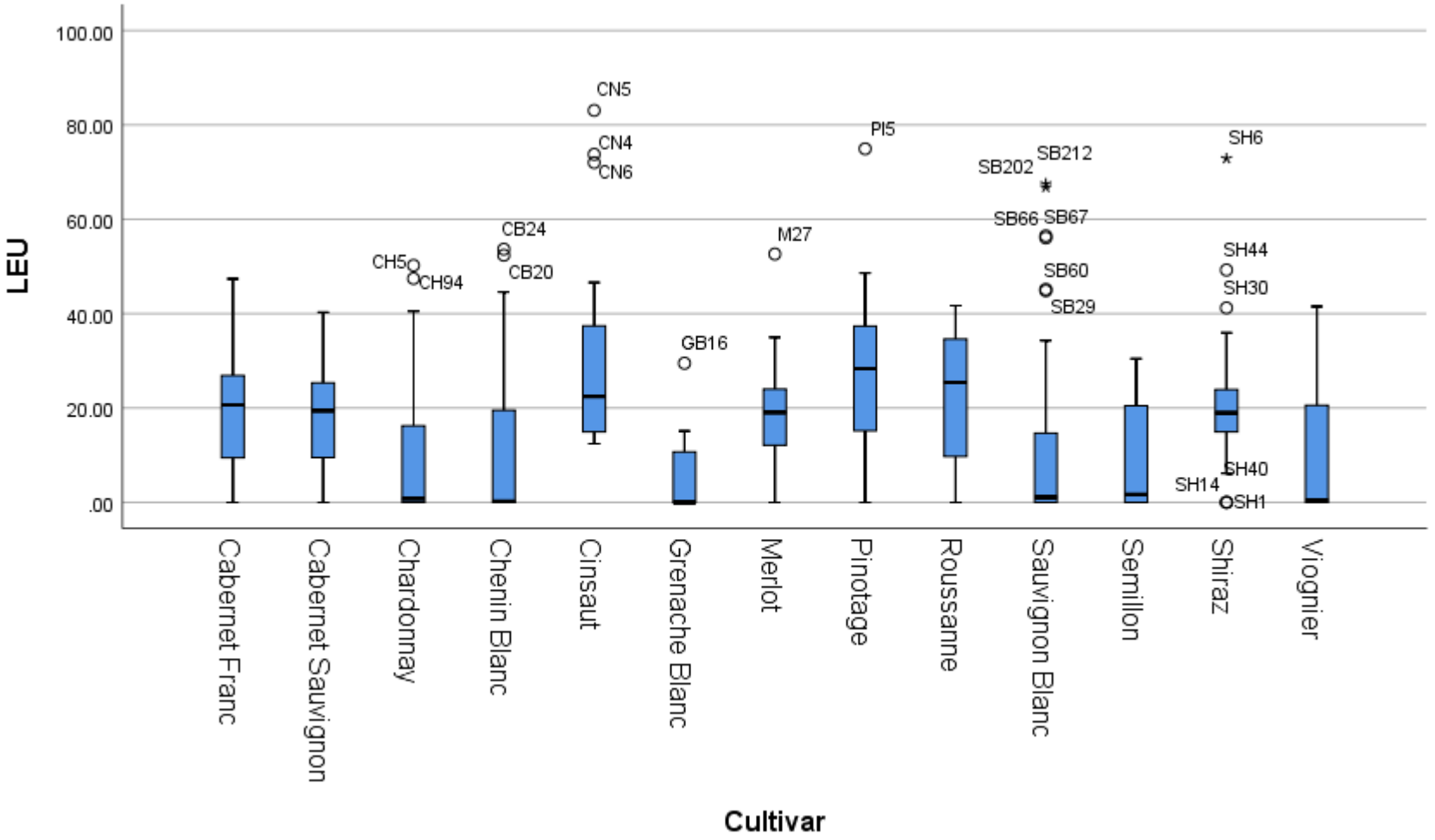


Figure B5.13. Box plots of leucine concentrations per cultivar (mg/L).

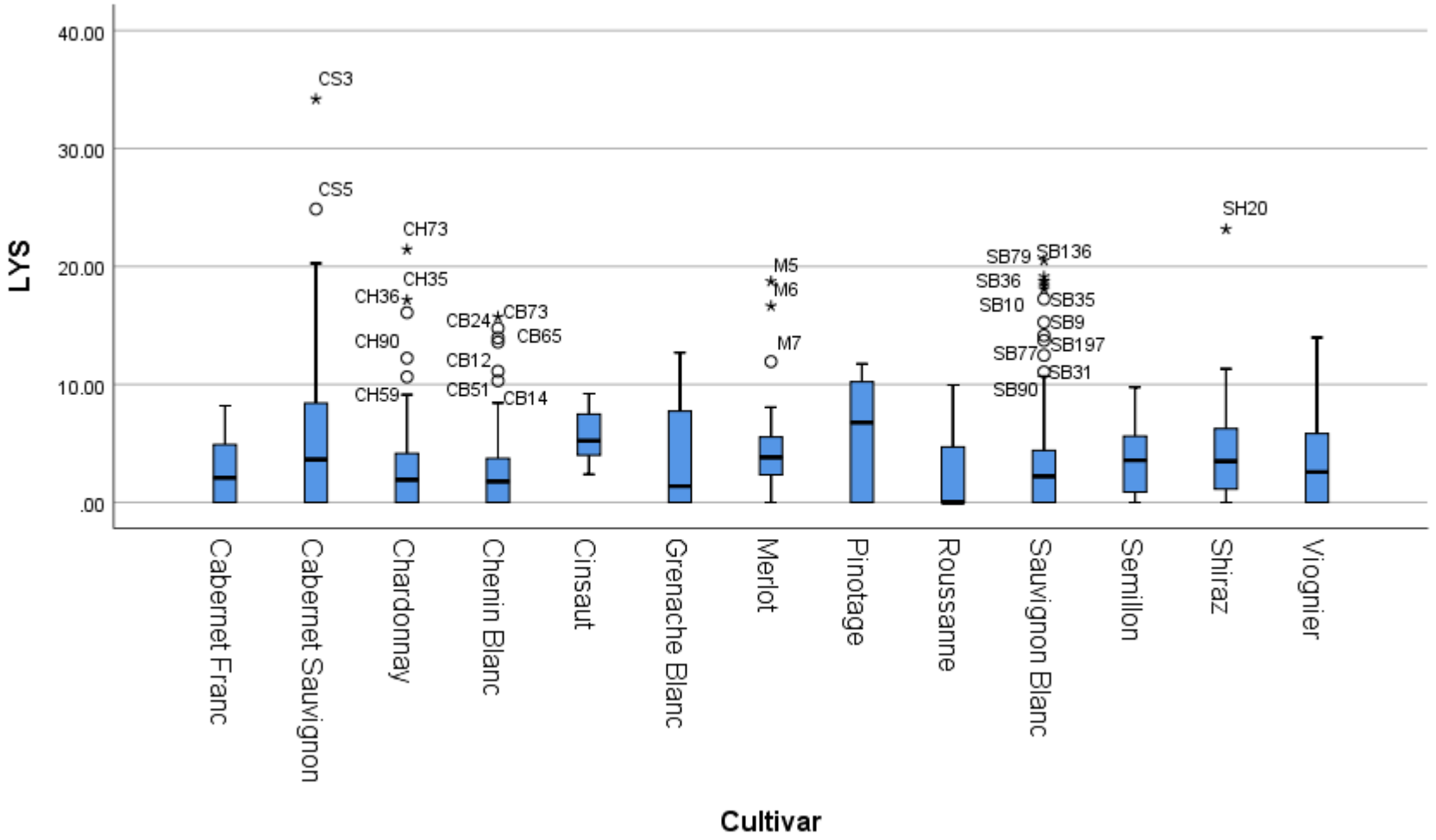


Figure B5.14. Box plots of lysine concentrations per cultivar (mg/L).

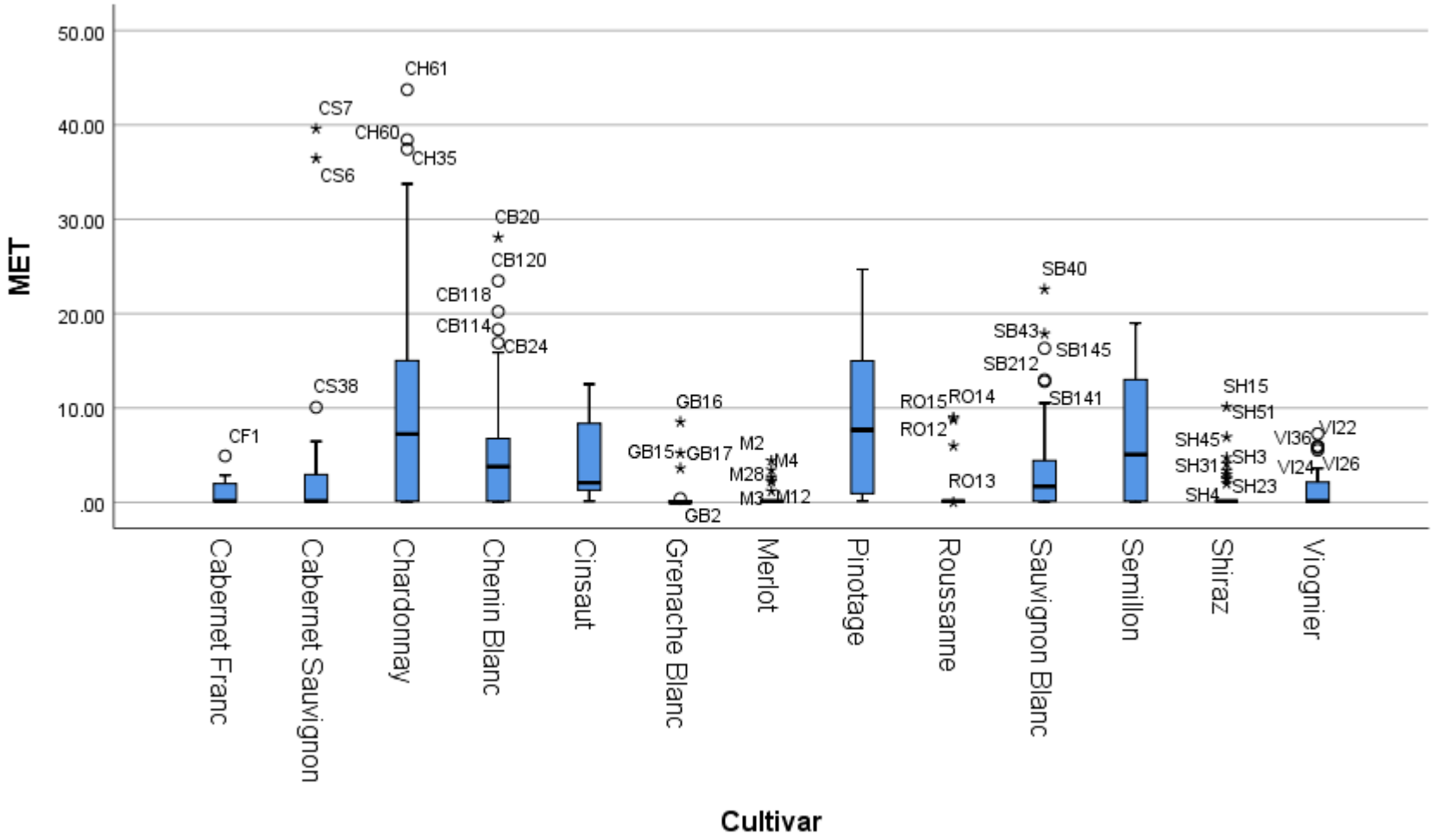


Figure B5.15. Box plots of methionine concentrations per cultivar (mg/L).

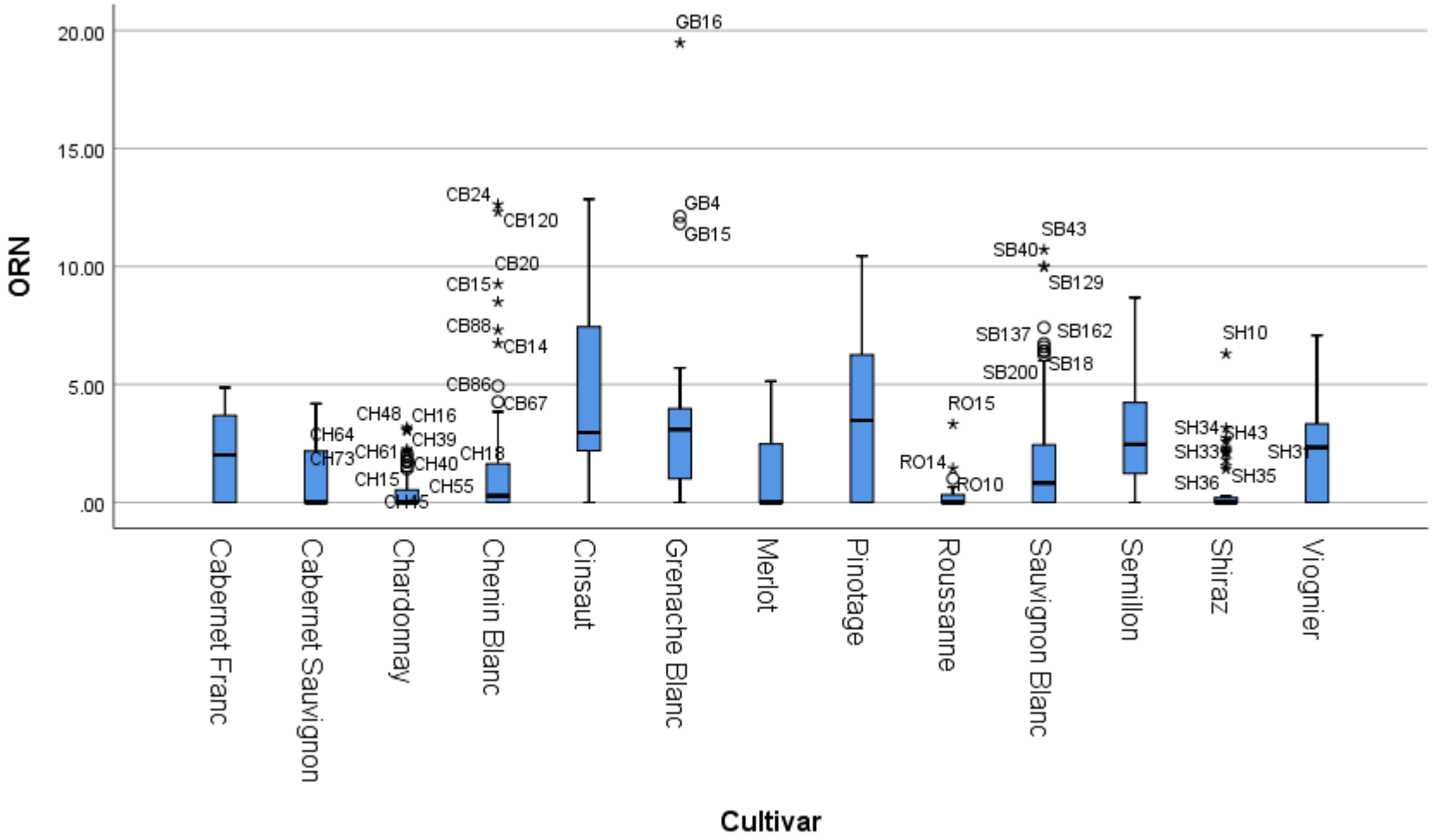


Figure B5.16. Box plots of alanine concentrations per cultivar (mg/L).

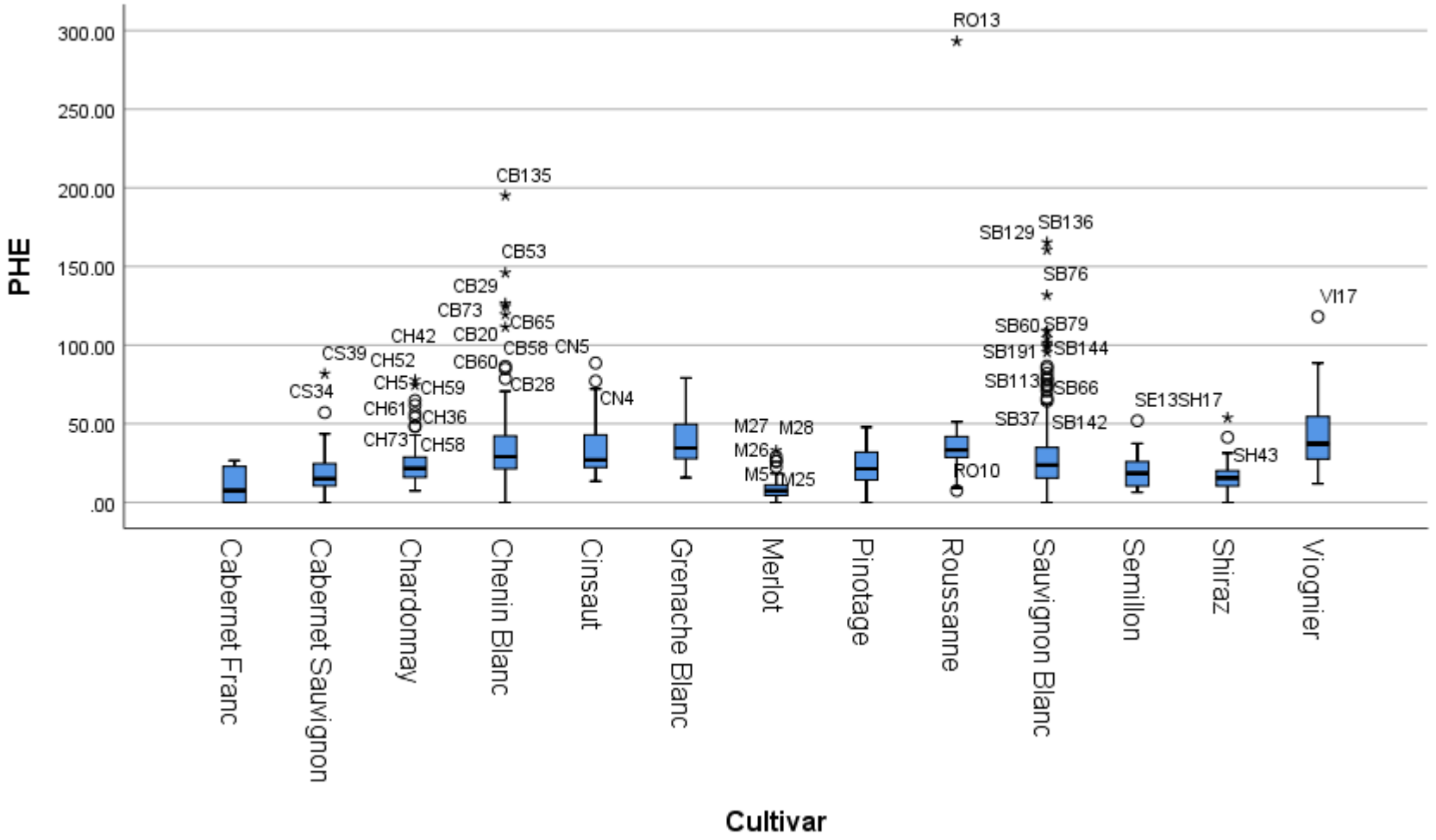


Figure B5.17. Box plots of phenylalanine concentrations per cultivar (mg/L).

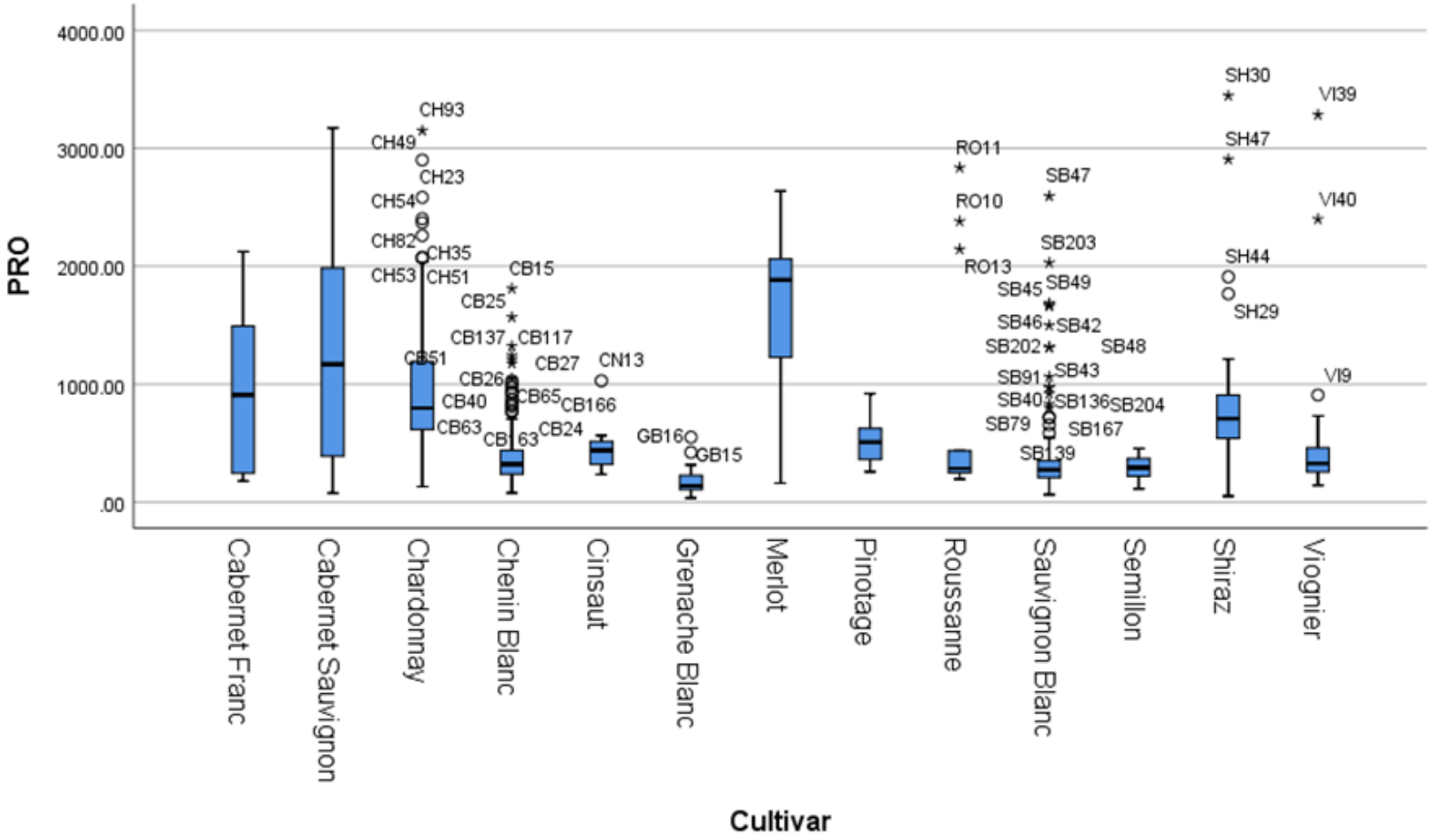


Figure B5.18. Box plots of proline concentrations per cultivar (mg/L).

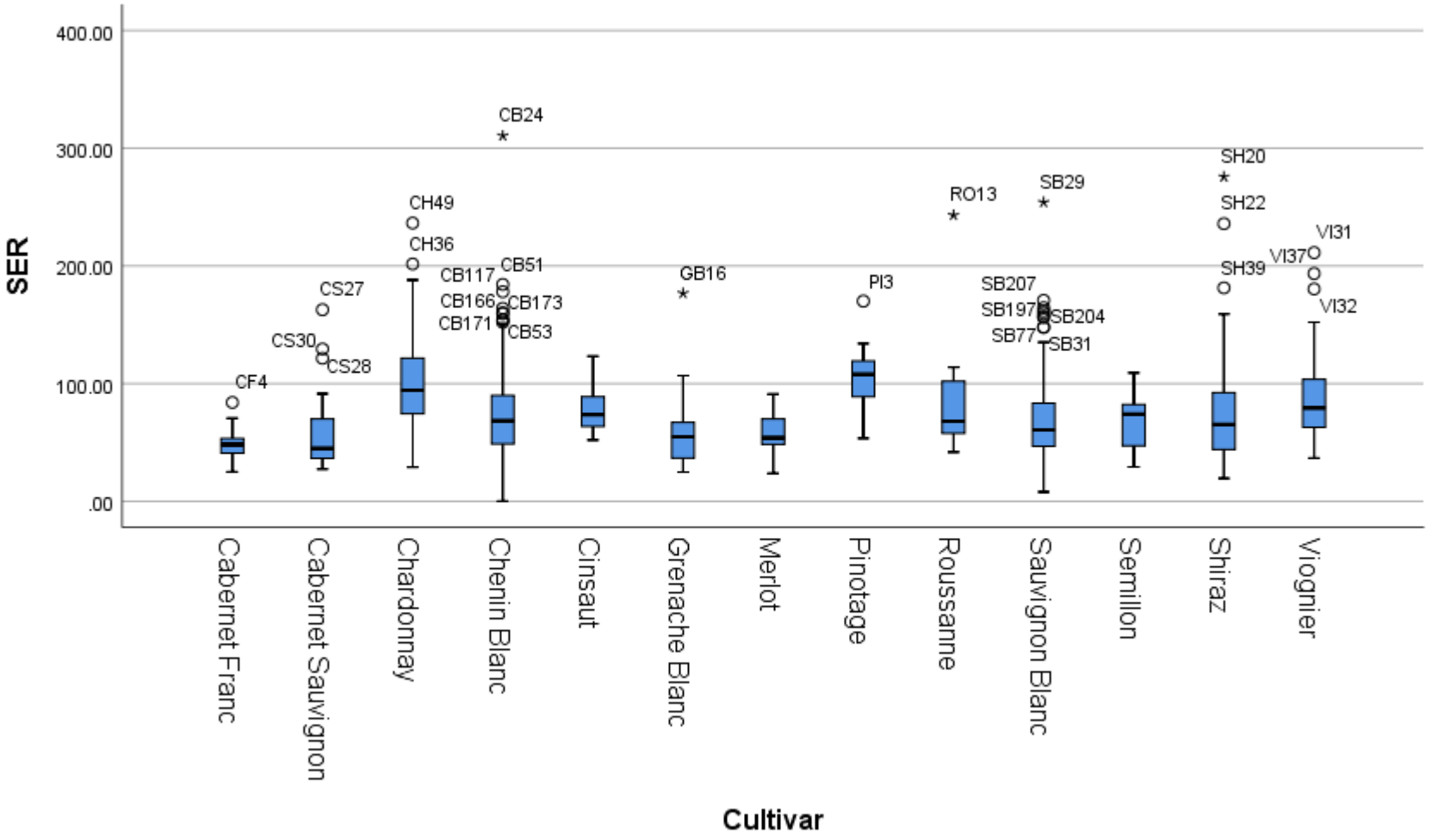


Figure B5.19. Box plots of serine concentrations per cultivar (mg/L).

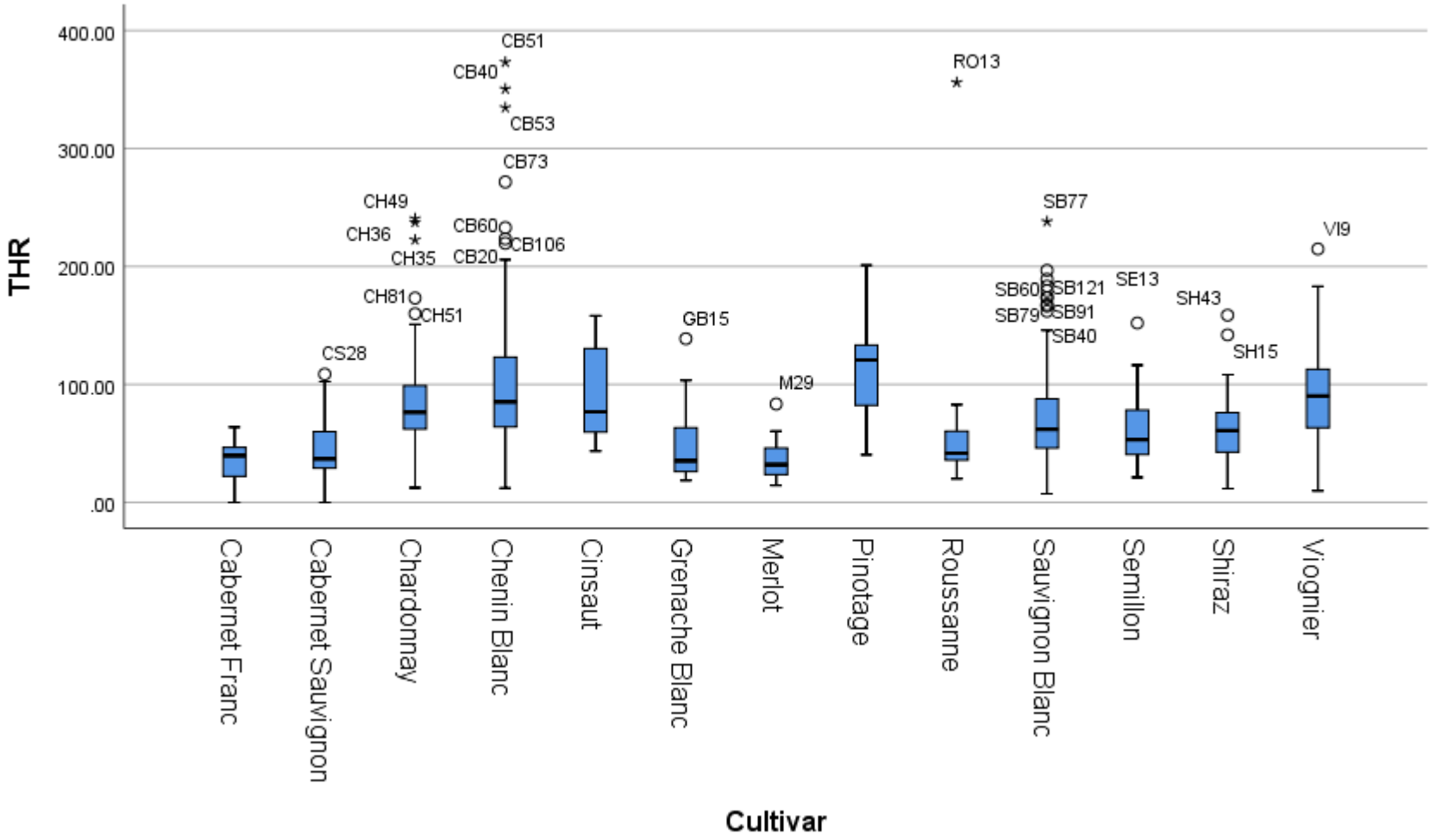


Figure B5.20. Box plots of threonine concentrations per cultivar (mg/L).

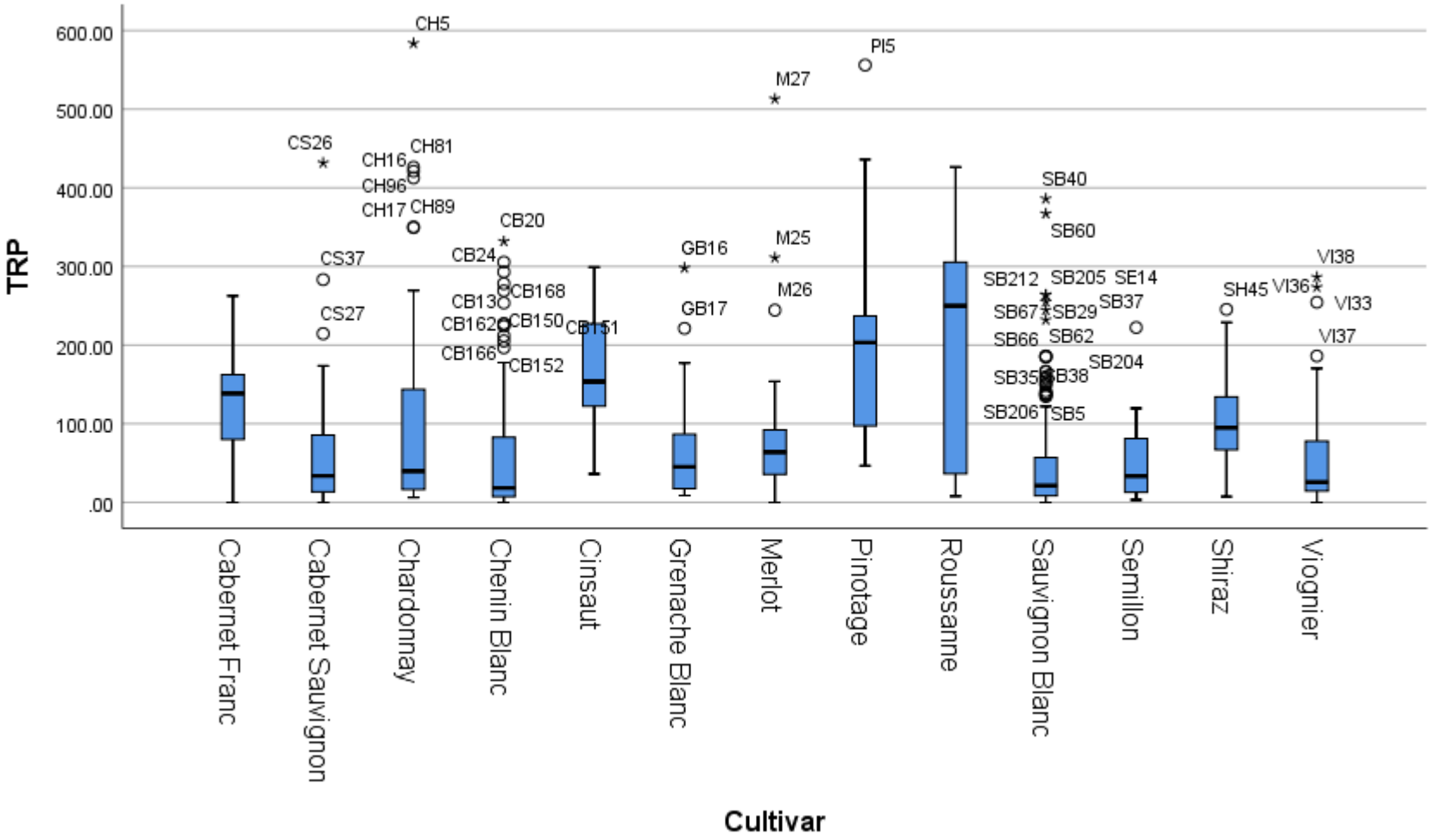


Figure B5.21. Box plots of tryptophan concentrations per cultivar (mg/L).

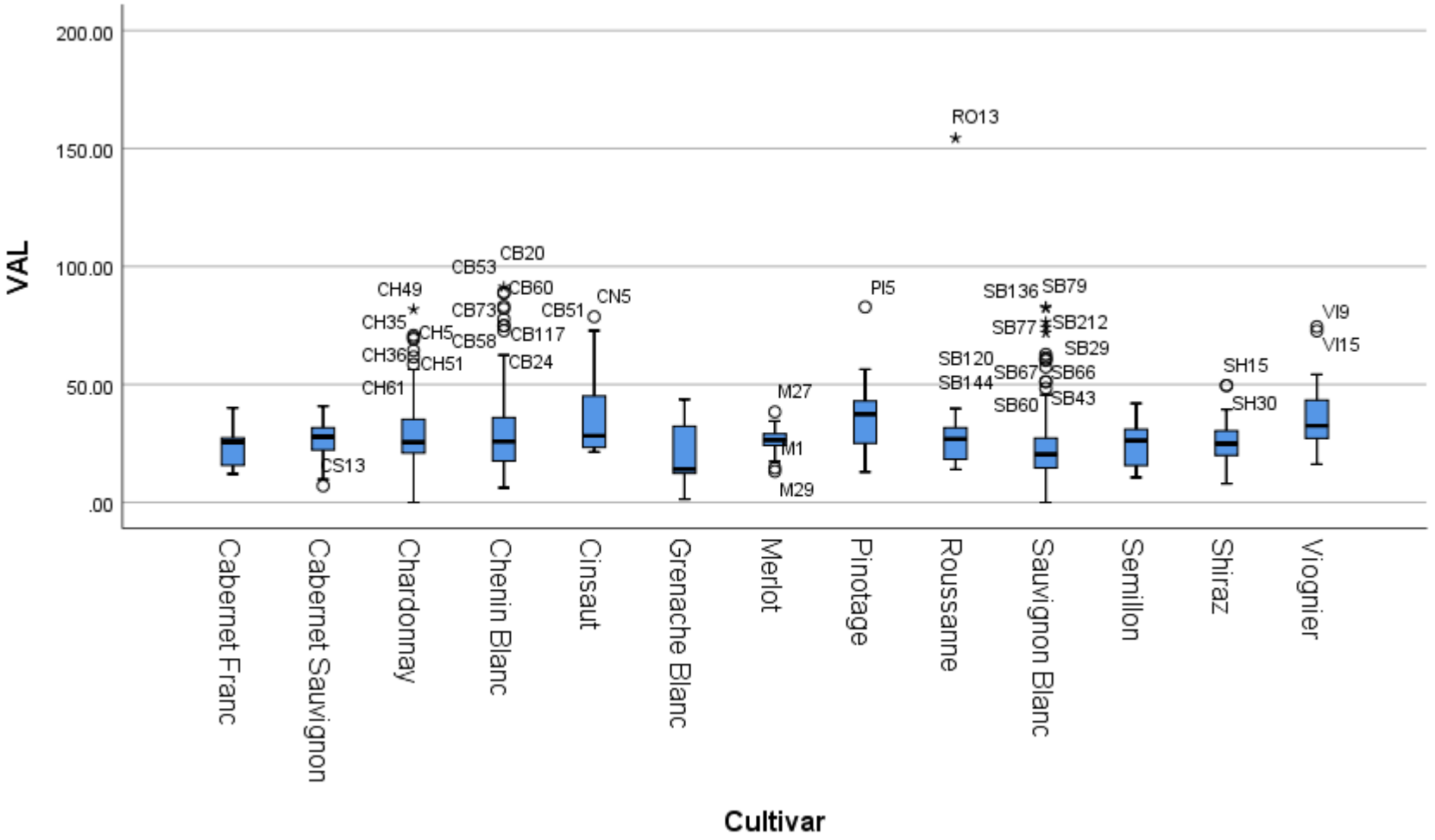


Figure B5.22. Box plots of valine concentrations per cultivar (mg/L).

Chapter 6

General Discussion and Conclusions

Chapter 6
General Discussion and Conclusions

Surveys on the YAN status (*i.e.* concentration and composition) of grape juices used for commercial winemaking have previously been conducted in various wine regions across the world (Kliewer, 1970; Huang & Ough, 1991; Spayd & Andersen-Bagge, 1996; Butzke, 1998; Stines *et al.*, 2000; Nicolini *et al.*, 2004; Hagen *et al.*, 2008; Nisbet *et al.*, 2014). However, no such study has yet been conducted in South Africa. This is of particular relevance due to the variability observed for YAN (as indicated by these surveys), and thus, the results obtained in a different region of the world may not be assumed true for South Africa. Therefore, the survey undertaken (Chapters 3 and 5) helped to gain insight into the YAN status of *local* grape juices used for commercial winemaking in South Africa. Furthermore, the unsupervised format aimed to gain an unimpeded view of the concentration and composition of YAN of the various cultivars relevant to the local wine industry. To this end, the results reported in Chapter 3 established the variability and range of total YAN concentrations for different cultivars grown in various districts across the Western Cape. This further helped to establish which cultivars in South Africa are most likely to require nutrient additions to ensure a successful fermentation, as well as those that could run the risk of excess nitrogen at the end of fermentation. The impact of geographical origin of the grapes was also explored and as a result, districts that may be frequently associated with nitrogen deficiency could also be identified. This information is invaluable to the local industry due to the current logistical issues associated with obtaining timely information regarding the nitrogen content of the grape juice matrix *before* the start of fermentation.

The *composition* of the nitrogen status of the grape juice matrix was also investigated taking into account the importance of the effect that YAN has on the aromatic profile of a wine (Hernández-Orte *et al.*, 2002; Torrea, 2003; Carrau *et al.*, 2005; Vilanova *et al.*, 2007; Mendes-Ferreira *et al.*, 2011; Torrea *et al.*, 2011; Barbosa *et al.*, 2012; Rollero *et al.*, 2018). Therefore, in addition to total YAN, the amount of nitrogen in inorganic (ammonium) and organic (amino acids) forms was also established – and found to vary between cultivars (Chapter 3). Furthermore, individual amino acid concentrations were also reported in Chapter 5 due to the roles they play in the complex metabolic activities of the yeast. This chapter helped to provide a comprehensive overview of the amino acid profiles of various industrially relevant cultivars by reporting on the most and least abundant amino acids, as well as the average proportion of various individual and groups (such as the branched-chain and aromatic) of amino acids. This work was done in hope of building a strong foundation of knowledge from which hypotheses can be generated for research on topics such as yeast nutrition and the impact of yeast metabolism on wine flavour and aroma. For example, due to the precursors provided by branched-chain and aromatic amino acids (Rapp & Versini, 1991), and other complex interactions of various amino acids with yeast metabolism, this research may help elucidate the

reason for certain aromas being associated with certain cultivars in addition to their characteristic varietal aromas. This is in line with work by Hernández-Orte (2002), who found that synthetic grape must solutions with amino acid profiles mimicking a specific grape variety resulted in the production of an aroma profile similar to that of a wine made with the actual grapes of the same variety.

For the aim of providing a more *comprehensive* understanding of such a complex and important component of the grape juice matrix, this data – originally collected for survey purposes – was mined for further value. Through exploratory data analysis techniques such as hierarchical clustering analysis and Classification and Regression Tree Analysis (CART), the important role that cultivar (and by extension, the genetics of the grapevine) plays in the resulting YAN profile, could be identified. These analyses revealed that cultivar outweighed both geographical origin as well as vintage in determining the concentration and composition of YAN. Moreover, the same data showed that cultivars that are more genetically closely related are more likely to have similar YAN profiles than those that are more distantly related. Following these findings, as well as those of Hernández-Orte (2002), the ability of the amino acid profile to discriminate between cultivars was tested using general discriminant analysis (GDA). Subsequently, using the best subset principle, amino acids that contributed to the best separation between cultivars could also be identified. These amino acids could serve as the basis for future work investigating varietal differences in aroma profiles and how different concentrations and ratios of amino acids affect the metabolic activities of yeast during fermentation. In addition, the discriminatory power of proline and arginine as proposed by Huang and Ough (1991) was tested. Although these amino acids were found to be significantly different between cultivars, they were not able to successfully discriminate between cultivars on their own.

In light of these findings, it is clear that a large amount of data can help identify underlying patterns, and, subsequently, the major factors that are at play. Thus, the *value* that could be obtained from this data was primarily due to the *volume* of data that was collected. '*Volume'* is particularly important in the context of YAN. This is strongly linked to the number of factors affecting its concentration and composition (Bell & Henschke, 2005) and, consequently, to the variability associated with this important component of the grape juice matrix. As highlighted in Chapter 2, a 'Big Data' approach to wine research, and specifically YAN, would be ideal to enable a holistic and integrated understanding of such a complex system. However, by reviewing the characteristics of Big Data, it became clear that the current *velocity* of YAN data generation (through traditional methods), may not be adequate to allow for a realistic 'Big Data' approach. Therefore, to facilitate further value creation, this study set out to set up a method which will enable high-velocity data generation.

Due to the simple, rapid, and cost-effective nature of spectroscopy and the recent developments in IR instrumentation and chemometrics, the ability of IR technology to accurately measure grape juice YAN was investigated (Chapter 4). Through the (i) unsupervised collection of a large number of samples from an array of different cultivars, districts, and vintages and (ii) the application of proper

external validation strategies, this study aimed to address the shortcomings of previous work seeking to calibrate IR instruments for the quantification of YAN. The sampling approach ensured a representative sample set for both the calibration and validation sets. Furthermore, due to the various validation strategies employed, it was clear that the proposed models would be capable of providing accurate results in a practical scenario where samples from different cultivars, vintages, and origins need to be analyzed. Therefore, this research provides not only a technique for effective Big Data collection but also a more rapid and cost-effective method for winemakers to obtain timely information. Looking at the number of parameters that can already be measured using IR spectroscopy, a wealth of information can be obtained from a single scan – an indispensable feature to both industry and research. Therefore, from a 'Big Data' point of view, IR spectroscopy is capable of providing *value* by means of collecting a high *volume* of a *variety* of data at a high *velocity*. Future success of this technology in the context of Big Data will be spurred on by the development of accurate calibrations on portable hand-held devices providing the means of on-line and real-time data collection.

The next step that would enable a more comprehensive understanding of this field of wine research would be the calibration of individual amino acids or certain relevant groups of amino acids. These groups include branched-chain and aromatic amino acids, which serve as precursors for higher alcohols, esters, and volatile acids, as well as sulfur-containing amino acids such as cysteine and methionine – said to play a role in H_2S and volatile thiol production. Unfortunately, due to the current limitations of the reference method, amino acids such as cysteine and tyrosine could not be accurately quantified. Nevertheless, future improvements on the separation and quantification of these amino acids, and subsequent IR calibrations will provide winemakers and researchers alike with an indispensable tool for more informed decision making.

Through the developments in IR spectroscopy and the collaborative effort to collect Big Data, the possibility of repeatedly producing quality wines at a standard that will keep up with consumer demands is fast becoming a reality, waiting to be taken advantage of.

References

Barbosa, C., Mendes-Faia, A., Mendes-Ferreira, A., 2012. The nitrogen source impacts major volatile compounds released by Saccharomyces cerevisiae during alcoholic fermentation. International Journal of Food Microbiology 160(2), 87–93.

Bell, S.-J. & Henschke, P.A., 2005. Implications of nitrogen nutrition for grapes, fermentation and wine. Australian Journal of Grape and Wine Research 11(3), 242–295.

Butzke, C.E., 1998. Survey of yeast assimilable nitrogen status in musts from California, Oregon, and Washington. American Journal of Enology and Viticulture 49(2), 220–224.

Carrau, F.M., Medina, K., Boido, E., Farina, L., Gaggero, C., Dellacassa, E., Versini, G., Henschke, P.A., 2005. De novo synthesis of monoterpenes by Saccharomyces cerevisiae wine yeasts FEMS Microbiol Lett 243(1), 107–115.

Hagen, K.M., Keller, M., Edwards, C.G., 2008. Survey of Biotin, Pantothenic acid, and Assimilable Nitrogen in Winegrapes from the Pacific Northwest. Am. J. Enol. Vitic. 59(4), 432–436.

Hernández-Orte, P., Cacho, J.F., Ferreira, V., 2002. Relationship between Varietal Amino Acid Profile of Grapes and Wine Aromatic Composition. Experiments with Model Solutions and Chemometric Study. Journal of Agricultural and Food Chemistry 50(10), 2891–2899.

Huang, Z. & Ough, C.S., 1991. Amino Acid Profiles of Commercial Grape Juices and Wines. Am. J. Enol. Vitic. 42(3), 261–267.

Kliewer, W.M., 1970. Free Amino Acids and Other Nitrogenous Fractions in Wine Grapes. Journal of Food Science 35(1), 17–21.

Mendes-Ferreira, A., Barbosa, C., Lage, P., Mendes-Faia, A., 2011. The Impact of Nitrogen on Yeast Fermentation and Wine Quality. Ciência Téc. Vitiv. 26(1), 17-32.

Nicolini, G., Larcher, R., et al., 2004. Status of yeast assimilable nitrogen in Italian grape musts and effects of variety, ripening and vintage. Vitis 43 (2), 89–96.

Nisbet, M.A., Martinson, T.E., Mansfield, A.K., 2014. Accumulation and Prediction of Yeast Assimilable Nitrogen in New York Wine-grape Cultivars. American Journal of Enology and Viticulture 65(3), 325–332.

Rapp, A. & Versini, G., 1991. Influence of nitrogen compounds in grapes on aroma compounds of wines In: Developments in Food Science 37. Elsevier 1659–1694.

Rollero, S., Bloem, A., Ortiz-Julien, A., Camarasa, C., Divol, B., 2018. Fermentation performances and aroma production of non-conventional wine yeasts are influenced by nitrogen preferences. FEMS Yeast Research 18(5), 1-11.

Spayd, S.E. & Andersen-Bagge, J., 1996. Free Amino Acid Composition of Grape Juice From 12 Vitis vinifera Cultivars in Washington. Am. J. Enol. Vitic. 47(4), 389–402.

Stines, A.P., Grubb, J., Gockowiak, H., Henschke, P.A., HøJ, P.B., Heeswijck, R., 2000. Proline and arginine accumulation in developing berries of Vitis vinifera L. in Australian vineyards: Influence of vine cultivar, berry maturity and tissue type. Australian Journal of Grape and Wine Research 6(2), 150–158.

Torrea, D., 2003. Production of volatile compounds in the fermentation of chardonnay musts inoculated with two strains of Saccharomyces cerevisiae with different nitrogen demands. Food Control 14(8), 565–571.

Torrea, D., Varela, C., Ugliano M., Ancin-Azpilicueta C., Leigh Francis I., Henschke P.A., 2011. Comparison of inorganic and organic nitrogen supplementation of grape juice – Effect on volatile composition and aroma profile of a Chardonnay wine fermented with Saccharomyces cerevisiae yeast. Food Chemistry 127(3), 1072–1083.

Vilanova, M., Ugliano, M., Varela, C., Siebert, T., Pretorius, I.S., Henschke, P.A., 2007. Assimilable nitrogen utilisation and production of volatile and non-volatile compounds in chemically defined medium by Saccharomyces cerevisiae wine yeasts. Applied Microbiology and Biotechnology 77(1), 145–157.